

**Three-Dimensional Shape from Shading:
Perception and Mechanisms**

Thesis by

Jennifer Y. Sun

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1996

(Submitted May 15, 1996)

© 1996

Jennifer Y. Sun

All Rights Reserved

Acknowledgements

I would like to express my deepest gratitude to my advisor, Professor Pietro Perona, without whom my graduate career could not possibly have been the relative smooth sailing it was. His insightful guidance and advice have made substantial differences in my work, as well as in my life. I would also like to thank the other members of my committee, Professors Bela Julesz, John Allman, Christof Koch, and Mark Konishi for their support. An entire chapter of this thesis would not have happened without the help of my collaborators, Andrea Mennucci and Luis Goncalves. I have benefited greatly from discussions with many colleagues and friends, most specially Dr. Achim Braun. I want to thank Robert Freeman, our wizard system administrator, who has been an irreplaceable resource. I would also like to express my thanks to Professors Richard Murray, Scott Fraser, and Steve Mayo for allowing me time on their computers.

My friends have been a crucial source of support throughout this entire process: my roommate Nicole for putting up with me for five unbelievable years, Bassil for helping me out numerous times on technical problems and otherwise, Lavonne for her support at work and at play, Brandon for sticking with me through my ups and downs, Doug for all the time we spent together, and Stefano for seeing me through to the very end.

And of course, none of this could have happened without my parents and their unflagging love and faith in me.

for my parents

Abstract

In this thesis, we address the issue of 3-D shape from shading by investigating shape perception in humans and the early vision mechanisms that subserve this perception. We first investigated the influence of scale, contour, and reflectance function on shape perception from shading. Our results suggest that subjects can form robust 3-D shape percepts that remain consistent across sittings for shapes of various contours and reflectance functions. We have found that salient 3-D percepts can be formed at the level of early vision mechanisms. Experiments in which a single target pattern is discriminated from multiple background distractors show that certain shaded, 2-D stimuli consistent with a top-lit, convex interpretation can be processed fast (<80 msec) and in parallel. Strong pop-out asymmetries and control experiments involving shaded patterns that do not have familiar 3-D interpretations suggest that such fast, parallel processing is indeed dependent upon perception of 3-D shape. We find that these mechanisms proceed most readily when the stimuli can be interpreted as convex and lit from top-left. These preferences for shape and lighting directions appear to be intrinsic to early vision and cannot be overturned using stereo disparity cues. These early vision 3-D mechanisms can also be influenced by 3-D contextual information. We report that, together with 3-D shape, apparent reflectance is computed fast as well. Moreover, it is apparent reflectance, rather than brightness or perceptual 3-D shape, that is the primary basis for discrimination during the early stages of visual processing.

Contents

Acknowledgements	iii
Abstract	v
1 Introduction	1
2 Measuring the Shape Percept	5
2.1 Experimental Methods	6
2.1.1 Method 1	6
2.1.2 Method 2	6
2.2 Subject Consistency	10
2.2.1 Tilt and Slant	10
2.2.2 Reflectance Function Effects	16
2.3 Perception of Shape	18
2.3.1 Overall Shape	19
2.3.2 Kurtosis	19
2.4 Tangent Plane Discontinuities	21
2.5 Scale Effects	21
2.6 Discussion	29
3 Processing of Shaded Patterns	32
3.1 Methods	33
3.1.1 Subjects	33

3.1.2	Apparatus and Stimuli	34
3.1.3	Procedure	34
3.1.4	Data Analysis	35
3.2	Experiments	36
3.2.1	Experiment 1 - Shaded Cubes vs Line Patterns	36
3.2.2	Experiment 2 - Shaded 2-D Patterns	39
3.2.3	Experiment 3 - Shaded Y-Junctions	40
3.2.4	Experiment 4 - Pop-out Asymmetries	43
3.2.5	Experiment 5 - Y-Junction in Circles	45
3.3	Discussion	46
3.3.1	Parallel Processing of 3-D Shapes	46
3.3.2	Shading Stimuli vs Line Stimuli	48
3.3.3	The Shaded Y-Junction	49
3.3.4	A Convex-lit-from-above Detector	50
4	Computation of Shape and Reflectance in Early Vision	52
4.1	Methods	56
4.2	Experiments	57
4.2.1	Experiment 1 - Shaded Triptych Patterns	57
4.2.2	Experiment 2 - Shaded Corner Patterns	58
4.2.3	Experiment 3 - Parallel or Serial?	62
4.3	Discussion	64
5	Direction of Lighting - Where is the Sun?	69
5.1	Methods	70
5.1.1	Method 1	70

5.1.2	Method 2	70
5.2	Experiments	71
5.2.1	Experiment 1 - Shaded Corners	71
5.2.2	Experiment 2 - Shaded Bubbles	74
5.3	Probable Causes	77
5.3.1	Lighting Preference and Handedness	82
5.3.2	Brain Lateralization and Lighting Preference	86
5.4	Discussion	89
6	Stereo and Shading	92
6.1	Methods	94
6.2	Experiments	94
6.2.1	Experiment 1 - Convex Shapes	94
6.2.2	Experiment 2 - Concave Shapes	96
6.3	Discussion	98
7	Influence of Contextual Information	101
7.1	Methods	101
7.1.1	Experimental Set-Up	101
7.1.2	Data Analysis	102
7.2	Experiments	102
7.2.1	Experiment 1 - Normal Orientation Cubes in Room Context .	102
7.2.2	Experiment 2 - Reverse Orientation Cubes in Room Context .	104
7.2.3	Experiment 3 - Holes in Walls	106
7.3	Discussion	108
7.3.1	Textural Effects	108

7.3.2	Expectation Effects	109
8	Conclusion	114
8.1	Summary of Results	114
8.2	The Biology of Shape from Shading	118
9	Appendix	123
9.1	A - The Surface Function	123
9.2	B - Method of Estimating Depth Scaling	124
9.3	C - Method of Estimating Psychometric Curves	124
	Bibliography	127

List of Figures

1.1	From shading cues alone we can extract a variety of shape information that can lead to a compelling sense of 3-D shape.	2
2.1	Stimulus screens 1, 2, & 3 are shown (left) with their respective xz-plane profiles (right).	7
2.2	Stimulus screens 4, 5, & 6 are shown (left) with their respective xz-plane profiles (right).	8
2.3	The elliptical gauge method for rendering 3-D shape perception into components of tilt and slant	9
2.4	For the first experimental paradigm in Method 2, during the first screen, a shaded bubble of systematically varied size was shown (top row). On the second screen, a pair of comparison curves were displayed (bottom row). The size of the comparison curves is determined by the angle they subtended, with the chord length scaled to match the diameter of the shaded bubble. For the second paradigm in Method 2, shaded bubbles (top row) were shown during both screens.	11
2.5	Each measured surface normal vector (solid) and its corresponding mean normal vector (dotted) were rotated together along the mean normal vector's line of constant tilt. When the mean normal vector lies on the xy-plane, the deviation in tilt between the mean normal vector and the measured normal vector corresponds to the actual angle between the vectors, irrespective of the original <i>slant</i> values.	12

2.6	Deviations from mean surface normals for two subjects and two surfaces are shown in scatter plots (top row), and standard deviations are binned according to slant values (bottom row).	14
2.7	Standard deviations of multiple subjects are shown here for Surface 1 (top) and Surface 4 (bottom). Standard deviations for tilt and slant are comparable, at around 5 degrees, for both surfaces.	15
2.8	Scatter plots of deviations in tilt and slant of Materials 2, 3, & 4 from Material 1 are shown for 2 subjects.	17
2.9	Surface plots of the real surfaces are shown in the left column. The middle and right columns show surfaces reconstructed from the data of two subjects for each of Surfaces 1, 2, & 3.	20
2.10	Apparent surface normals (solid) for 2 subjects each are shown here for Surfaces 4, 5, & 6 in comparison with the actual surface normals (dotted). The solid line indicates the locali of tangent plane discontinuities for the surface.	22
2.11	Influence of image size on perceived depth scaling is shown here for 6 subjects.	23
2.12	Percentage of time a subject judged the bubble to be less shapely than the comparison arcs is plotted in A against the angle subtended by the comparison arcs. B shows the same psychometric curves with fitted error functions superimposed.	24
2.13	The means obtained from fitting error functions to psychometric curves are plotted here with dotted lines against bubble size in degrees of visual arc. Fitted lines are plotted in solid, with the slopes as indicated.	25

2.14	Graph A shows the percentage of time that subject ST judged the shaded bubble shown first, a 1-degree bubble, to be more curved than a second shaded bubble, which could be 1, 3, or 5-degrees. In B, the first bubble shown was a 3-degree bubble, and in C, a 5-degree bubble.	26
2.15	Graph A shows the percentage of time that subject ST judged the shaded bubble shown first, a 0.5-degree bubble, to be more curved than a second shaded bubble, which could be 0.5, 2, or 5-degrees. In B, the first bubble shown was a 2-degree bubble, and in C, a 5-degree bubble.	27
2.16	Graph A shows the percentage of time that subject EM judged the shaded bubble shown first, a 1-degree bubble, to be more curved than a second shaded bubble, which could be 1, 3, or 5-degrees. In B, the first bubble shown was a 3-degree bubble, and in C, a 5-degree bubble.	28
3.1	Top row shows a sample test screen (A) and a sample mask screen from the shaded cubes experiment (B). Bottom row shows sample test screens from the line cubes experiment (C) and the line Y-junctions experiment (D).	37
3.2	SOA for 75% accuracy for 3, 12, and 24 display sizes are shown for each of the three patterns: the shaded cubes, the line cubes, and the line Y-junctions.	38
3.3	SOA for 75% accuracy for three shaded control patterns, shaded tiles (A), shaded X-junctions (B), and shaded T-junctions (C), are shown in comparison with that of the shaded cubes.	41

3.4	The graph on the right summarizes subjects' performance for this hexagonal shaded pattern (left). The performance is parallel for target-absent trials.	42
3.5	Stimulus display screen from the pyramids experiment is shown on the left, and the results are shown on the adjacent graph (right).	43
3.6	A stimulus display screen from the shaded Y-junction in circles experiment (left) and the corresponding performance (right)	44
3.7	"Reverse" orientation shaded cubes are shown on the left. Results from this experiment is shown on the graph at right.	45
3.8	"Reverse" orientation shaded pies are shown on the left. Results from this experiment is shown on the graph at right.	46
3.9	Experiments were conducted on gapped Y-junction in circles of two sizes (top). The corresponding performance graph is shown below each pattern.	47
4.1	Pop-out behavior is seen when the task is to spot an inverted target among upright cubes. When the target and distractor patterns are reversed, serial search behavior is seen.	53
4.2	According to the Feature Detection Theory proposed by Treisman & collaborators, pop-out occurs when the target, but not the distractors, is recognized by the feature detector.	54

- 4.3 The perception of shaded patterns may be represented with a shape map and a reflectance map. If the shaded pattern leads to the perception of a 3-D shape, then the apparent differences in luminance are perceived as shading, and the resulting apparent reflectance is uniform. If the perception is of a flat surface, then the apparent luminance differences are attributed to changes in reflectance. 55
- 4.4 The right eye's view of Conditions 1-4 in Experiment 1 is shown in A. B shows data collected from 5 naive subjects performing at a display duration of 120 msec. Significant impairment of performance is seen only when the 3-D shape was perceived as single-reflectance, and the luminance pattern can be interpreted as shading (Condition 1). . . . 59
- 4.5 Cross-fusing stereo pairs for Condition 1 (A) and Condition 2 (B) of Experiment 1 are shown. (C) Performance averaged over 5 naive subjects for Conditions 1, 2, & 3 (both distractors and target were flat) indicate a marked asymmetry between Conditions 1 & 2. (D) Wire-frame patterns were used as controls. The direction of asymmetry is opposite to that of the shaded conditions. Display duration was 250 msec. 61
- 4.6 (A) Stimulus patterns from Conditions 1 & 2 of Experiment 1 were tested in two display sizes. When the display size was increased from 12 to 24, the 75% accuracy display duration increased significantly for Condition 1 but not for Condition 2. (B) For Conditions 1 & 2 in Experiment 2, when the display size was increased from 12 to 24 items, there was a significant increase in the 75% accuracy display duration for Condition 2 but not for Condition 1. 63

- 4.7 Upside-down cubes, forming a “C” pattern, are embedded among upright ones. All the upright cubes are identical in luminance pattern. The darkest side of the upside-down cubes, however, varies from light to dark, while the other two sides are held constant. Which of the upside-down cubes has a luminance pattern that is identical to the upright ones? After you have made your choice, turn the figure upside-down and make your choice again. Do your choices agree? 65
- 4.8 (A) We suggest that the top-lit, upright cube pattern promotes 3-D shape perception by early vision processes more so than the upside-down one. Such a difference in shape perception would imply a difference in reflectance perception. (B) A multiple-reflectance patch would pop-out from a background of single reflectance, while a single-reflectance patch would be difficult to spot in a multiple-reflectance background. (C) Stimulus displays that have reflectance maps equivalent to those on the left, but no perceptual 3-D shape, result in a similar asymmetry in performance under controlled experimental conditions. Data was taken from 2 naive subjects at a display duration of 120 msec. 67
- 4.9 This diagram summarizes schematically our proposed scheme for 3-D pop-out. While we do not completely rule out pop-out arising directly from 3-D shape, our results suggest that this is weak compared to the pop-out that arises from reflectance differences. 68
- 5.1 Shaded bubbles are typically interpreted as 3-D shape with the assumption that lighting is coming from above. 70

5.2	An example of a 0-degree stimulus screen is shown in A. Lighting from the left is denoted by positive angles from the vertical; a 30-degree screen is shown in B, and a 90-degree screen is shown in C. D shows a right-lit screen(-60 degrees).	72
5.3	Results for 6 conditions displayed at a duration of 180 msec are shown. "A" conditions have distractors that are lit from the left, and "B" conditions have distractors that are lit from the right.	73
5.4	Lighting orientation, as defined by the distractors, is plotted on the x-axis; negative degrees indicate lighting from the right and positive degrees indicate lighting from the left. The SOA duration estimated for 67% accuracy is plotted on the y-axis.	75
5.5	Each bar gives the average difference across 12 subjects between the performance for a right-lit condition with its corresponding left-lit condition. Results are given for the 30, 60, and 90-degree pairs. The p-value for significance is given above each bar.	76
5.6	The working environment of a right-hander is usually set up with the light on the left.	79
5.7	Because most students are right-handed, classrooms tend to be set up with windows on the left.	80
5.8	Another biological asymmetry we have is our asymmetry in brain lateralization. Most people are right-brain lateralized for spatial-perceptual tasks, while some are flipped with respect to the norm.	81
5.9	The predictions that follow from each hypothesis are summarized in this chart.	82

5.10	A parabolic curve is fitted to the relevant portion of the data of each subject. This graph gives as an example the data and the fitted curve for Subject AW. For AW, the preferred angle of lighting is 16 degrees left of the vertical.	83
5.11	Preferred angle of lighting was averaged across right-handed subjects and left-handed subjects separately.	84
5.12	Preferred angle of lighting is plotted against a handedness score for each subject. The handedness score is derived from the Oldfield (1977) Inventory, a standardized questionnaire for evaluating handedness. . .	85
5.13	Handedness is plotted on the x-axis, and difference in performance between target present in left-hemifield trials and target present in right-hemifield trials is plotted on the y-axis.	87
5.14	Preferred direction of lighting is plotted against hemispheric dominance for our 3-D pop-out task, as determined by hemifield preference . . .	88
5.15	The lighting directions of paintings from 3 major museums are determined by two naive subjects. Results show a marked tendency for artists to choose top-left lighting.	90
5.16	3-D icons and logos found on the net are almost exclusively rendered with top-left lighting.	91
6.1	(A) Displays similar to those shown for Conditions 1 & 2 were shown at a duration of 130 msec. An asymmetry in performance is observed (B). Conditions 3 & 4, displayed at 200 msec, also resulted in an asymmetry (C), but in the opposite direction.	93

6.2	When 3-D shape is conferred using stereo disparity cues, the asymmetry in performance is enhanced for both the normal cubes (display duration = 130 msec) and the inverted cubes (200 msec).	95
6.3	When concave 3-D shape is conferred using stereo disparity cues, the asymmetry is not significantly enhanced for the normal cubes (130 msec), and is significantly reduced for the inverted cubes (200 msec).	97
7.1	Performance of three subjects was tested using shaded cubes of different sizes (left). The results are shown in comparison to that of same-size shaded cubes on the right.	103
7.2	Top row shows normal orientation cubes arranged in floor perspective (left) and normal orientation cubes in a room context as well as in floor perspective (right). Bottom row shows the respective effects of these enhancements in terms of improvement. These plots reflect the data collected from 4 and 3 subjects, respectively.	104
7.3	Normal orientation cubes were arranged in ceiling perspective and room context (left). The improvement of this condition over the standard task is shown on the right.	105
7.4	Reverse orientation cubes arranged in floor perspective (left) and with room context (right) is shown on the top row. Bottom graphs depict the respective effects.	106
7.5	Reverse orientation cubes were also tested in ceiling perspective (left). The results are summarized in this graph in terms of improvement (right).	107

7.6	Consistent distractor holes in 3-D wall (A), inconsistent distractor holes in 3-D wall (B), and frame with no 3-D interpretation for control experiments (C)	112
7.7	Improvement of performance for the consistent frame condition over the control condition is shown at left, and improvement of performance for the inconsistent frame condition is shown at right.	113
8.1	A schematic summarizing our findings concerning the processing of 3-D shapes by early vision processes.	122

List of Tables

2.1	Reflectance Parameters of Materials	16
2.2	Systematic Deviations from Material 1	18

Chapter 1 Introduction

When we look at the world around us, we see it as three dimensional. No matter if we are viewing a physical 3-D scene, or even just a black-and-white photograph, our sense of shape is compelling. Even when stereo, color, and texture cues are absent, grey-level images nonetheless contain many cues from which we can build a 3-D percept — luminance edges, shading gradients, occlusion contours, cast shadows, to name a few (see Figure 1.1). Although the saliency of 3-D shape perception from shading has been noted and used by artists since at least the 4th Century B.C. (Gombrich, 1976), little yet is understood about how 3-D shape is actually represented in the brain.

The reason that research in shape from shading has been limited in the past may be attributed to the lack of sophisticated computer tools for generating realistic images and for collecting psychophysical data. Nonetheless, the existing literature reveals that our visual system can derive with ease many aspects of a 3-D scene from shading. (Rittenhouse, 1786; Brewster, 1847; Gibson, 1950; Horn, 1975; Yonas *et al.*, 1979; Pentland, 1982; Todd & Mingolla, 1983; Ramachandran, 1988; Mingolla & Todd, 1989).

One of the first issues to be studied in this field was the issue of measuring the subjective shape percept. Historically, the study of form or shape perception has dealt primarily with shapes of only two dimensions. Even when the third dimension was studied, the question was typically posed in terms of the perceived slant in depth of planar surfaces. As the sophistication of computer graphics capabilities grew, so did the body of work that addresses the issue of 3-D shape perception and representation

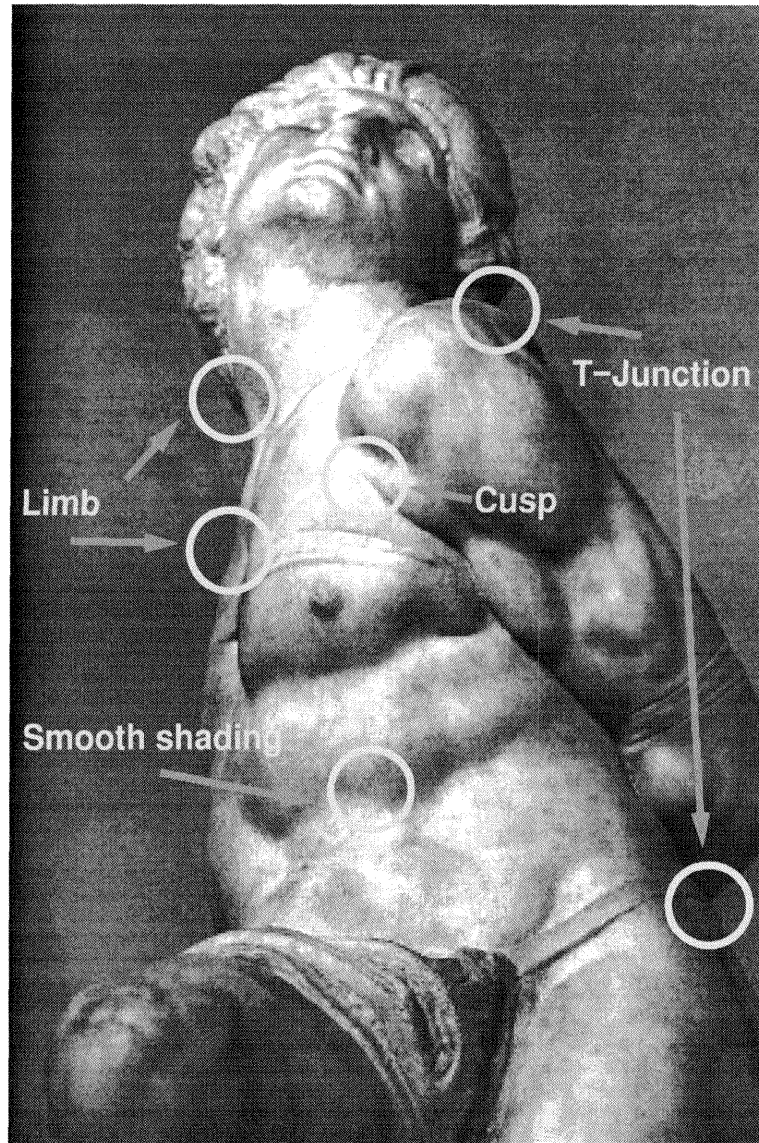


Figure 1.1: From shading cues alone we can extract a variety of shape information that can lead to a compelling sense of 3-D shape.

(Braunstein *et al.*, 1982; Cutting & Millard, 1984; Lappin *et al.*, 1980; Todd & Mingolla, 1983; Stevens & Brookes, 1987; Todd & Akerstrom, 1987; Todd & Reichel, 1989; Bulthoff & Mallott, 1992; Koenderink *et al.*, 1992).

In Chapter 2, we evaluate the most recently proposed method for quantifying our internal representation of a 3-D shape (Koenderink *et al.*, 1992). We will also address some of the primary questions regarding the representation of 3-D shape in the brain. For instance, how consistent is our perception of 3-D shape from shading? Does it change from viewing to viewing? How does changing the reflectance function or the size of the stimulus affect perception? And of course, how accurate are we in our perception?

A second major issue in the study of shape perception from shading concerns the mechanisms underlying the perception of 3-D shape from shading. We know that shading information can supply a wealth of cues from which we are able to compute 3-D shape. Yet which and how, and along what time course, are these cues combined in the process of 3-D perceptual build-up?

In studying visual processing, the first step is to examine the processes that occur in early vision. Early vision, otherwise known as preattentive vision, includes those mechanisms which underlie the initial stages of visual processing (See Neisser, 1967; Beck, 1982; Treisman, 1982; Julesz, 1984). These mechanisms are capable of operating in parallel across the visual field, and is believed to be capable of detecting simple features only, and not the conjunction of these simple features. Color, luminance, orientation, motion, and textons are some examples of such simple features. Since shaded stimuli that can be interpreted as 3-D shapes are typically conjunctions of luminance patches and oriented lines, one would not expect them to be readily processed by these early vision mechanisms.

During the last few years, however, there has been accumulating evidence suggesting that some aspects of 3-D shape perception are computed by early vision processes (Braun 1990, 1993; Enns & Rensink 1990, 1991; He & Nakayama 1992, 1993; Kleffner & Ramachandran 1992, Adelson 1993, Sun & Perona 1996a, 1996c). Experiments that demonstrate perceptual pop-out involving shaded patterns with familiar 3-D interpretations have provided particularly convincing evidence. In our studies, we investigate specifically this preattentive processing phase of 3-D perception.

We present our findings confirming that 3-D shape perception is indeed subserved by early vision processes in Chapter 3. In Chapter 4, we show that, despite the fact that both shape and reflectance are computed in parallel by early vision, reflectance, and not shape, is used as the primary cue by processes mediating segregation and pop-out.

In the remaining chapters of this thesis, we investigate in more detail various characteristics of early vision 3-D mechanisms. Chapter 5 explores the lighting-from-above assumption that seems to be crucial for fast and parallel processing of shaded stimuli as 3-D shapes. Chapter 6 describes experiments in which we explore the interaction between shape-from-stereo and shape-from-shading mechanisms, and in Chapter 7, we ask whether these early vision 3-D processes can be influenced by 3-D contextual cues.

Chapter 2 Measuring the Shape Percept

In studying shape perception, the first problem is in measuring and expressing quantitatively this subjective shape percept in a meaningful way. Because this is a non-trivial problem with no obvious solution, a variety of methods have been proposed (Stevens, 1983a, 1983b; Todd & Mingolla, 1983; Todd & Akerstrom, 1987; Mingolla & Todd, 1989; Todd & Reichel, 1989; Bulthoff & Mallott, 1992), most of which are well-suited for answering the specific question their designers have in mind, but not easily adaptable for other studies of shape perception. In their 1992 paper, Koenderink *et al.* proposed a new method for measuring subjective percept of a 3-D surface that seems to be more widely applicable. In our experiments, we used Koenderink's method to investigate the effect of various stimulus parameters on subjective 3-D shape perception from monocularly presented shaded stimuli. The viability of this method subject consistency across sittings and variations of reflectance function is addressed in Section 2.2. In Section 2.3 we look at inter-subject agreement and accuracy as compared to ground truth. The specific shape parameters whose perceptual effects we wish to address, 1) boundaries and contours and 2) stimulus size, will be discussed respectively in Sections 2.4 & 2.5. From studying these perceptual effects, we hope to attain some understanding of the underlying mechanisms and algorithms and their roles in shape perception from shading.

2.1 Experimental Methods

Two different experimental paradigms were used in this series of experiments. Method 1, based upon the method proposed by Koenderink *et al.* (1992), was used in Sections 2.2 to 2.5 for measuring subjective shape percept of shaded stimuli that spanned 5 or more degrees of visual angle. Method 2, a 2AFC paradigm, was used in Section 2.5 for stimuli spanning fewer than 5 degrees.

2.1.1 Method 1

6 surfaces based upon a unit disc on the frontoparallel xy-plane of the screen were rendered using the built-in graphics capabilities of a Silicon Graphics IRIS (See Figures 2.1 & 2.2, Appendix A for more details). We varied the reflectance function and size of these presented figures. Subjects viewed the images monocularly. For measuring subjective shape percept, a method proposed by Koenderink *et al.* (1992) was used (See Figure 2.3). Subjects adjusted a gauge figure controlled via the mouse to reflect the perceived local surface normal orientation at 69 different points on a given surface. 8 naive subjects were used for these experiments.

2.1.2 Method 2

To investigate specifically the effect of scale on perceived “shapeliness”, we presented linearly shaded bubbles, much like the ones used by Ramachandran (Ramachandran, 1988; Kleffner & Ramachandran, 1992), of different sizes and shading gradients monocularly on a constant background. Note that the stimuli are 2D patches that are shaded linearly, but look convincingly three dimensional. Subjects did not notice that they are not realistically shaded. The size and shading gradients were randomized,

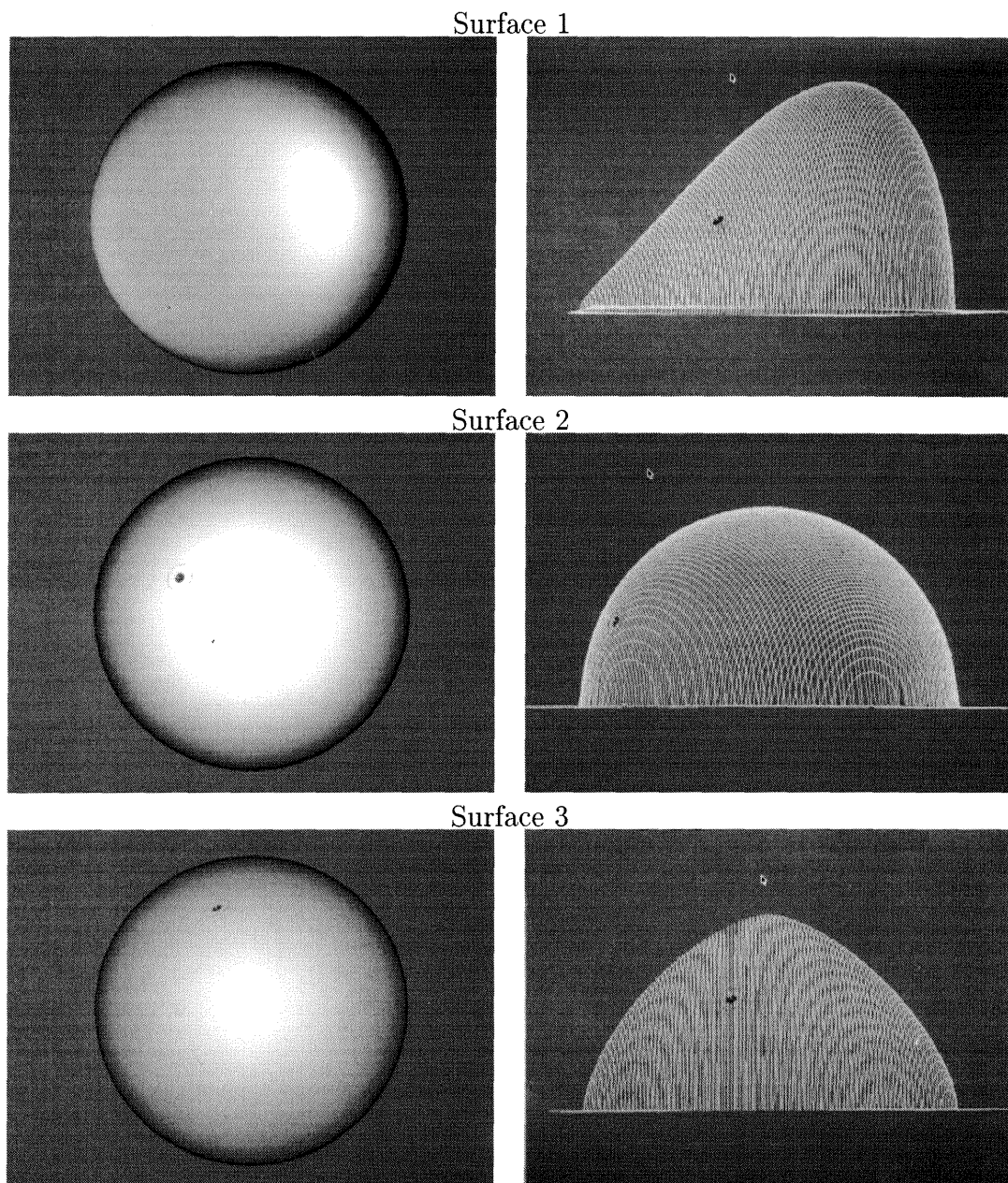


Figure 2.1: Stimulus screens 1, 2, & 3 are shown (left) with their respective xz-plane profiles (right).

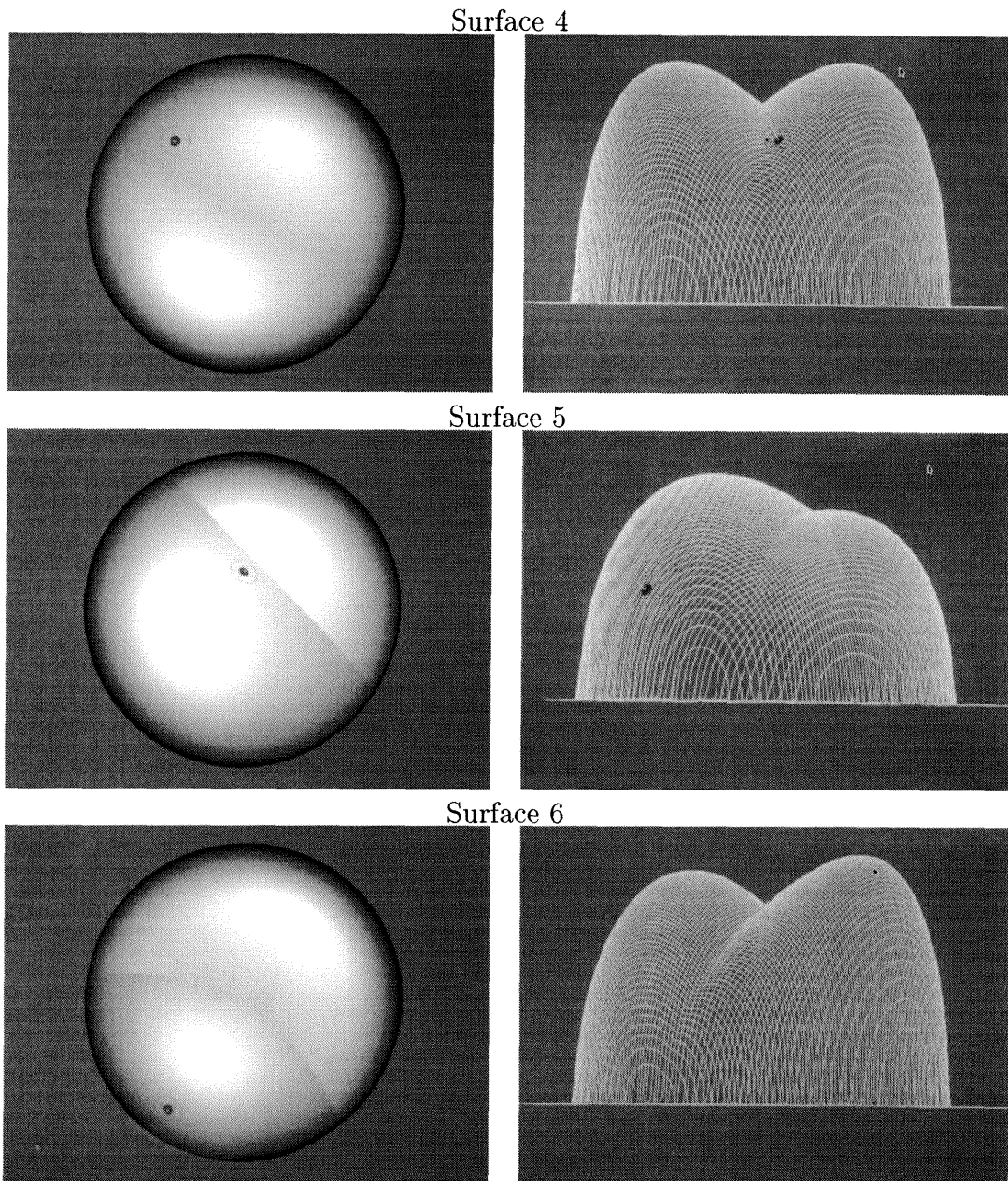
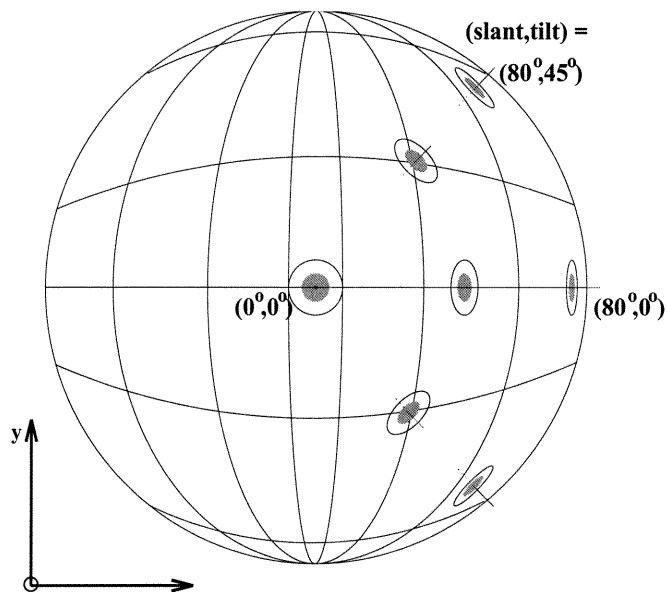
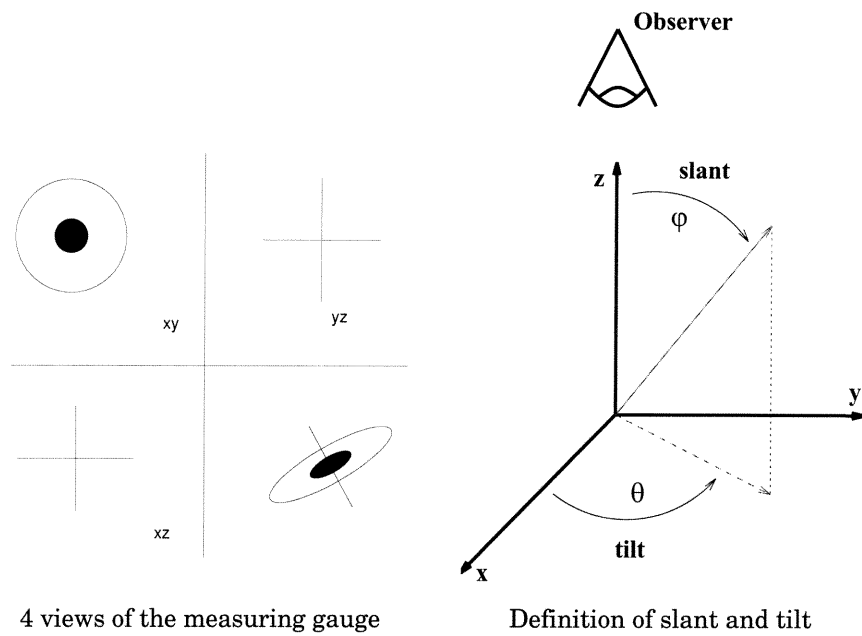


Figure 2.2: Stimulus screens 4, 5, & 6 are shown (left) with their respective xz-plane profiles (right).



Examples of surface slant and tilt measurements on a sphere

Figure 2.3: The elliptical gauge method for rendering 3-D shape perception into components of tilt and slant

and subjects were asked to evaluate in a 2AFC paradigm whether a shaded bubble presented in the first screen is more or less curved than 1) arcs of systematically varied length in degrees scaled so that chord length matched the diameter of the bubble, or 2) another shaded bubble, presented on a subsequent screen (See Figure 2.4). About 900 trials were presented in each block of experiments. 4 naive subjects were used for these experiments.

2.2 Subject Consistency

2.2.1 Tilt and Slant

We first investigated the reliability of measuring subjective perception using Method 1 by looking at intra-subject consistency. Data was collected using the same shapes from the same subjects on 3 or more sittings, much like the paradigm used by Koenderink *et al.* (1992).

We first calculated a mean surface normal for each one of the 69 points of a surface from the data generated from multiple sittings. The deviation in angle between each experimental surface normal and its mean was obtained after applying a coordinate transform that places the mean surface normal in the xy-plane (See Figure 2.5). This deviation was then decomposed into two components: one in the tilt direction and one in the slant direction. This method allows deviation magnitudes in the tilt direction to be independent of the slant value. If the deviation were calculated without the coordinate transform, when the surfaces normals have small slant values (ones corresponding to points close to local maxima), even very small errors will show up as large deviations in the tilt direction.

Figure 2.6 shows the range of deviations seen for two of our stimulus shapes,

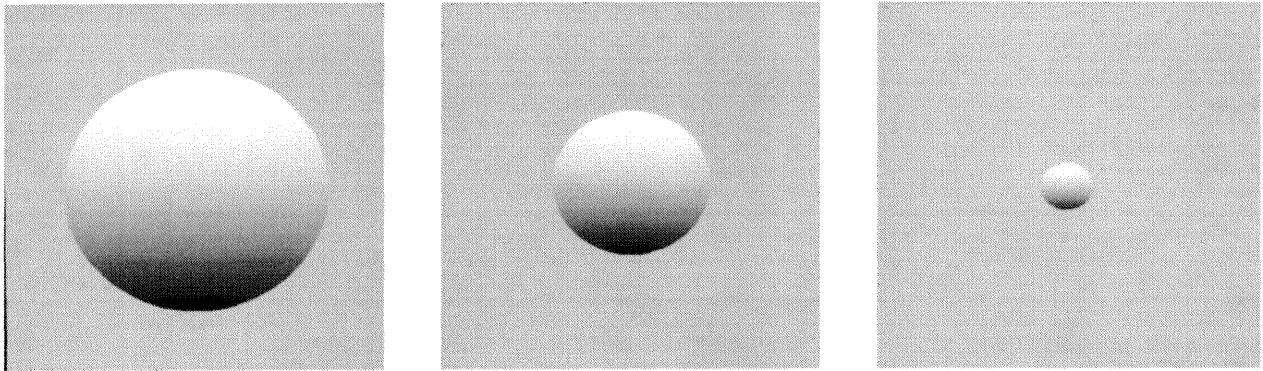
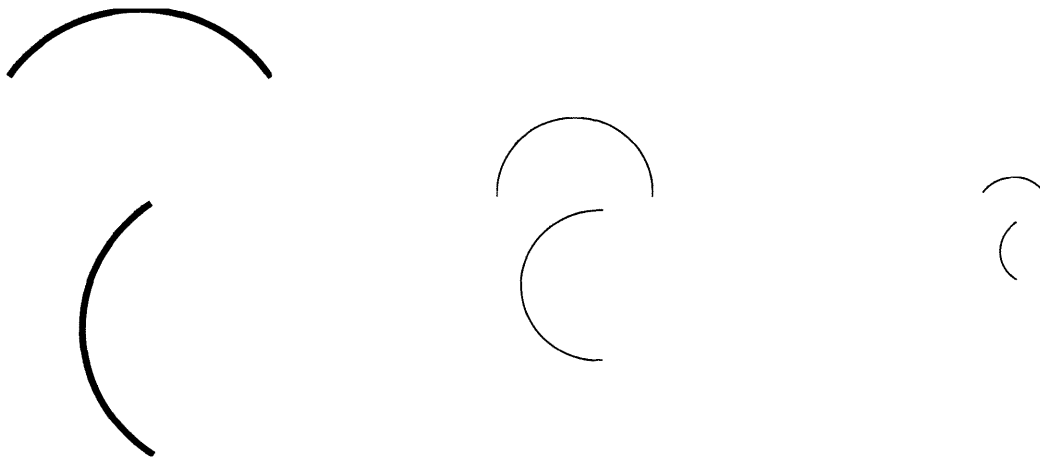
Screen 1**Screen 2**

Figure 2.4: For the first experimental paradigm in Method 2, during the first screen, a shaded bubble of systematically varied size was shown (top row). On the second screen, a pair of comparison curves were displayed (bottom row). The size of the comparison curves is determined by the angle they subtended, with the chord length scaled to match the diameter of the shaded bubble. For the second paradigm in Method 2, shaded bubbles (top row) were shown during both screens.

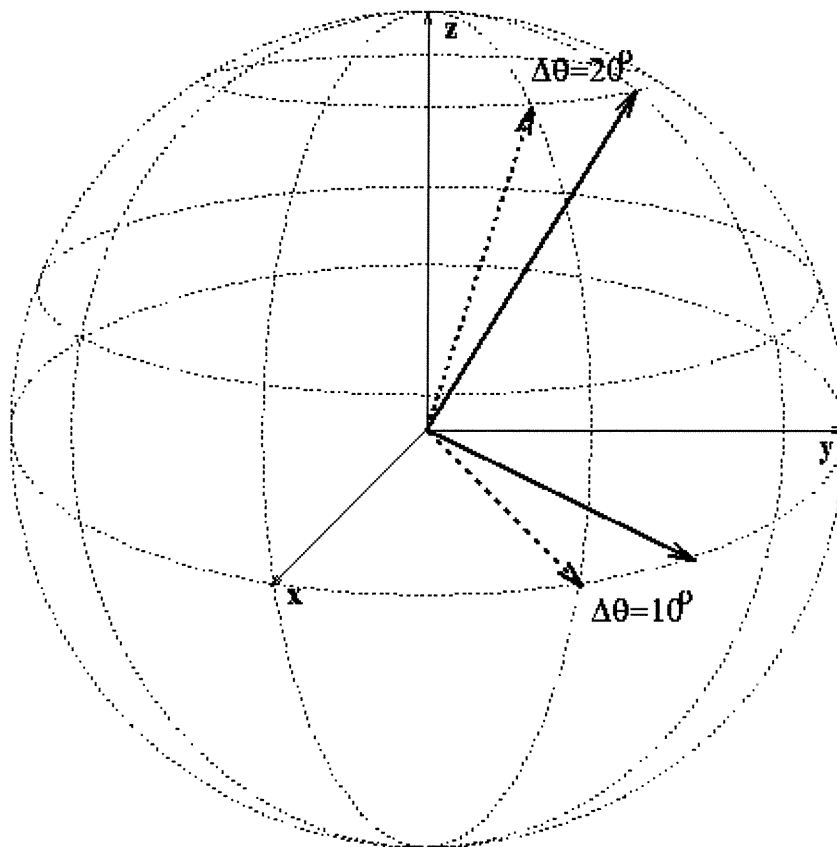


Figure 2.5: Each measured surface normal vector (solid) and its corresponding mean normal vector (dotted) were rotated together along the mean normal vector's line of constant tilt. When the mean normal vector lies on the xy -plane, the deviation in tilt between the mean normal vector and the measured normal vector corresponds to the actual angle between the vectors, irrespective of the original slant values.

Surface 1 and Surface 4. In Figure 2.6, the left panel shows scatter plots of deviations from two subjects, and the right panel shows standard deviations from means binned according to slant values. There appears to be a tendency towards higher precision as slant increases. Figure 2.7 summarizes consistency for 4 subjects for each of the two surfaces. As can be seen, subjects maintain good consistency from sitting to sitting. As reported by Koenderink *et al.* (1992), tilt consistency is very good. We found deviations to be generally around 5 degrees. However, while they reported slant judgment to be imprecise, much more uncertain than that of tilt, we found deviations in the slant direction to be comparable to those in the tilt direction, ranging between 5 and 10 degrees. Even for Surface 4, a relatively complicated shape with two maxima and a discontinuity, both tilt and slant measurements continue to have good consistency.

While not proof in and of itself, this high degree of intra-subject consistency we observe is consistent with and lends credence to the following hypotheses:

1. Subjects form a reliable 3-D impression of the shaded figures. This would further imply that whatever a priori model the brain uses to form this 3-D percept, it is a relatively fixed and stable model.
2. This method of measuring subjective perception is adequately intuitive and does not require the subject to express his 3D shape percept in quantities that are far removed from the psychophysical variables used by the visual system for encoding surface orientation.

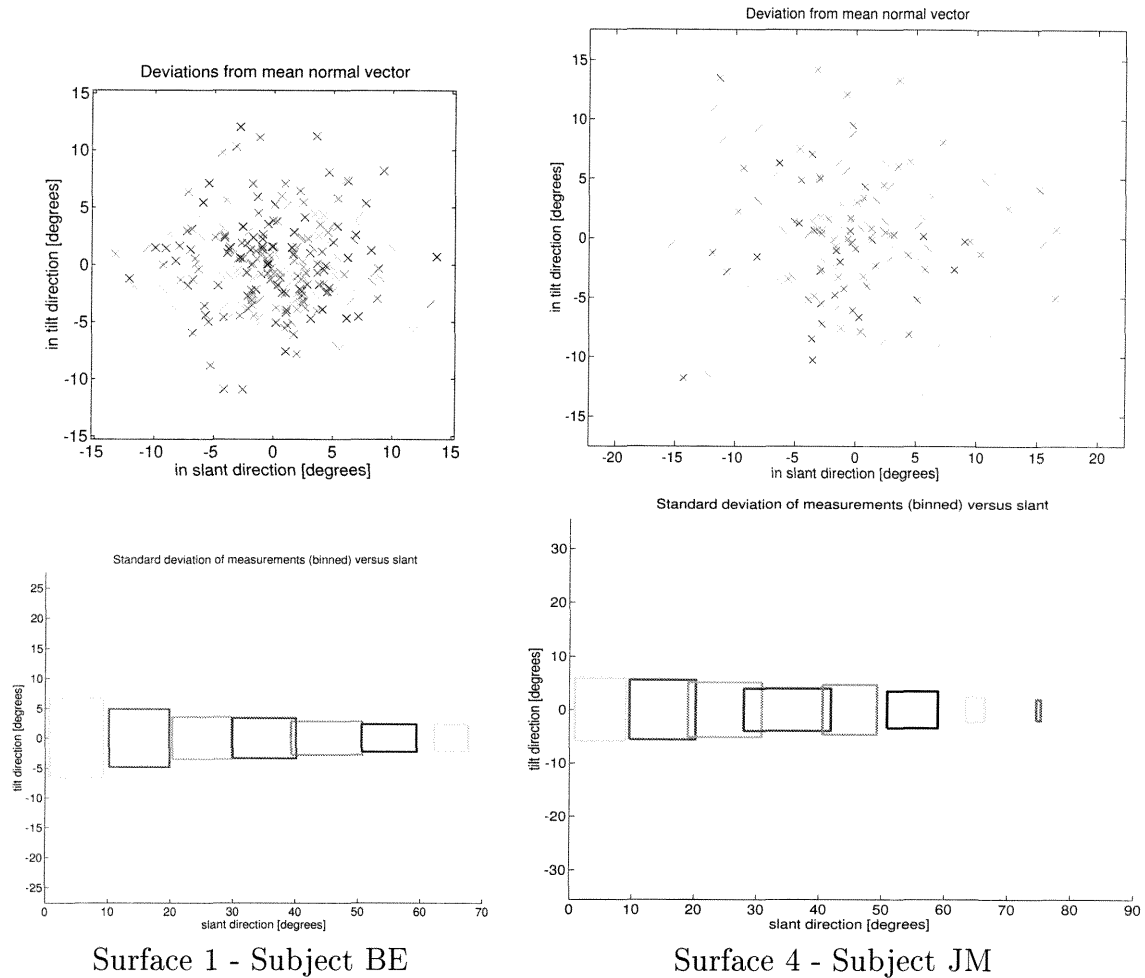


Figure 2.6: Deviations from mean surface normals for two subjects and two surfaces are shown in scatter plots (top row), and standard deviations are binned according to slant values (bottom row).

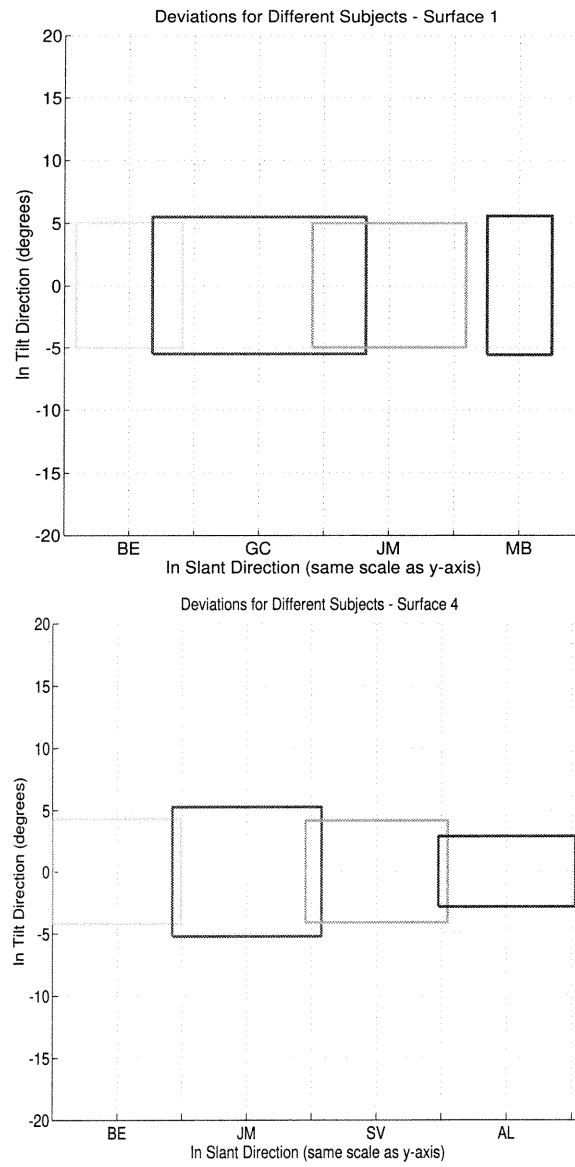


Figure 2.7: Standard deviations of multiple subjects are shown here for Surface 1 (top) and Surface 4 (bottom). Standard deviations for tilt and slant are comparable, at around 5 degrees, for both surfaces.

Materials	Diffuse Reflectance	Specular Reflectance	Specular Scattering
1	0.60	0.40	6.0
2	0.99	0.00	10.0
3	0.00	0.99	50.0
4	0.00	0.99	6.0

Table 2.1: Reflectance Parameters of Materials

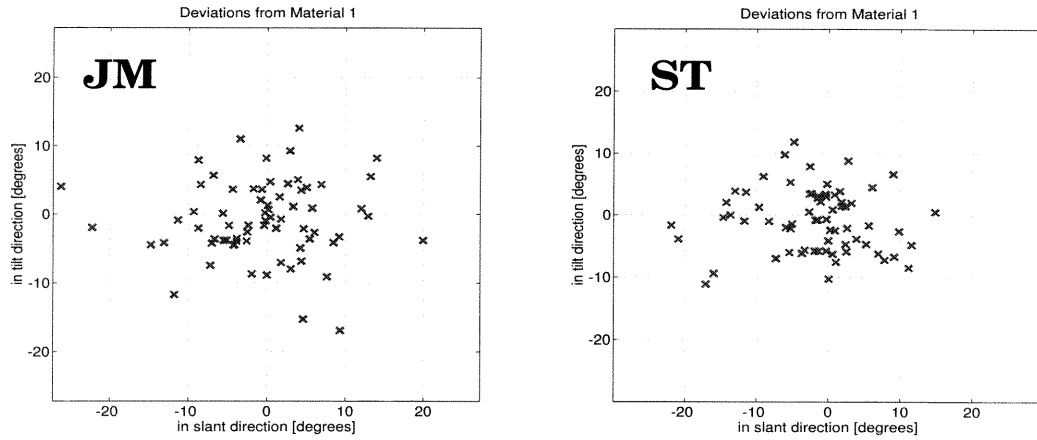
2.2.2 Reflectance Function Effects

The SGI graphics library allows one to define the "material" of an object by specifying parameters of diffuse reflectance from 0 to 1, specular reflectance from 0 to 1, and specular scattering exponent from 0 to 128. To investigate the effect of reflectance function on shape perception we chose 4 different materials that covered a wide range in each of the parameters:

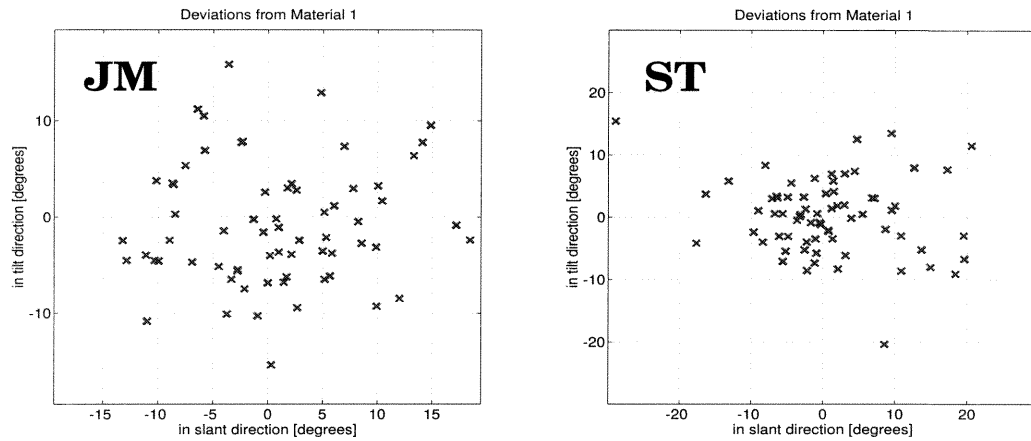
For each subject, we used the apparent surface normals collected using Material 1 as a baseline with which to compare the data collected with Materials 2, 3, & 4. The surface normal deviations between each of these materials and Mat 1 for two subjects are depicted in the scatter plots shown in Figure 2.8. In all cases, the differences between the Mat 2, 3, 4 surface normals and the Mat 1 surface normals lie within the range of deviations expected for data collected from different sittings (5 to 10 degrees).

If the different reflectance functions were to cause a change in the impression of shape or depth scaling, one would expect a systematic shift of the deviation distributions away from the origin of the graph. The plots in Figure 2.8 do not seem to show significant amounts of such shifts. One way of obtaining a numerical representation of these shifts would be to find the mean deviation from Mat 1 for each surface. Data from a perfect observer who perceives the surface of a test material as completely

Material 2



Material 3



Material 4

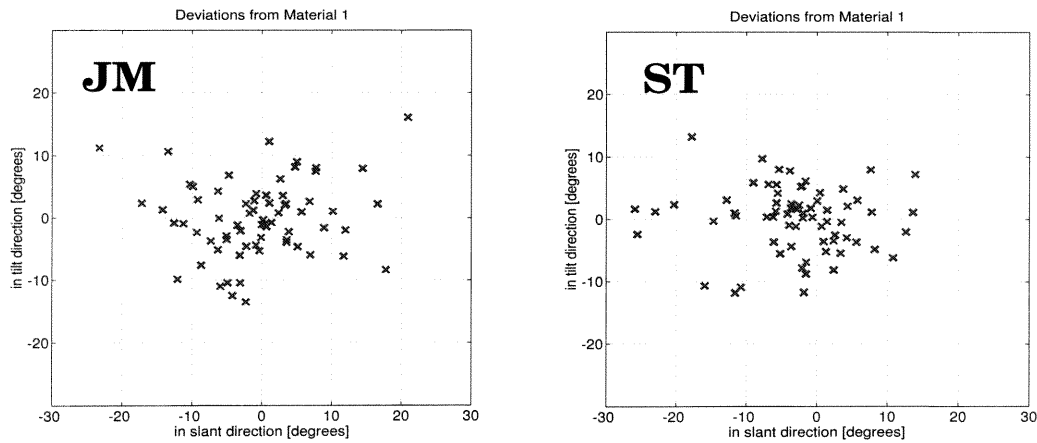


Figure 2.8: Scatter plots of deviations in tilt and slant of Materials 2, 3, & 4 from Material 1 are shown for 2 subjects.

		Material 2		Material 3		Material 4	
Subject	Surface	tilt	slant	tilt	slant	tilt	slant
MB	1			0.63	-0.70	0.86	-1.54
ST	1	-0.66	-2.39	0.21	0.80	-0.09	-3.93
ST	2	-0.32	-0.67				
JM	2			0.66	-2.90	-0.53	-1.46
MB	2	-0.22	-1.85				
JM	3	0.91	0.53	-0.98	0.61	-1.51	-0.11
average		-0.073	-1.095	0.130	-0.548	-0.318	-1.760

Table 2.2: Systematic Deviations from Material 1

identical to that of Mat 1 would give a mean deviation of zero. Table 2.2 shows the mean values in degrees for 3 real observers in the tilt and slant directions. Mat 4 resulted in the largest amount of shift in slant, averaging to almost 2 degrees between subjects, corresponding to perhaps a more flattened percept. However, since mean deviations of around 2 degrees are typically seen between different sittings even for the same material, this shift is not significant.

2.3 Perception of Shape

The results from the previous section suggest that subjects do indeed form a rather robust percept of 3-D shape from these shaded figures, one that is stable across different sittings and not affected very much by changes in reflectance function. This consistency is however not necessarily related at all to accuracy. In this section we take a look at how subjective perception compares to the ground truth.

2.3.1 Overall Shape

To obtain a general idea of what shape the subjects were perceiving, we asked the subjects to describe the overall shape of their percept for each experiment by drawing the perceived xz-plane profile of the stimulus surface, the same view as that shown in the right column of Figures 2.1 & 2.2, and indicating the relative location(s) and height(s) of the maxima. The apparent shapes reported were in all cases qualitatively similar to the test surfaces, even for the more complicated stimuli, Surfaces 4, 5, & 6.

2.3.2 Kurtosis

While the overall apparent shapes were qualitatively to similar to the real surface and therefore qualitatively similar between subjects, we noticed from subjective reports that different subjects experienced different impressions of depth, an observation noted also by Koenderink *et al.* (1992). We obtained a quantitative representation of this depth impression by assuming that the perceived surface is a scaled version in height of the real surface. We then found the scaling factor (k) that when multiplied with the real surface gives surface normals that fit the data best in a least squares manner (See Appendix B for more details). Figure 2.9 shows the reconstructed surfaces and scaling factors obtained for different subjects. We observed from our analysis that there is a pronounced tendency to underestimate the depth of the surface, and that this perceived depth varies from subject to subject. The surface plots shown were chosen to represent the extremes in each case.

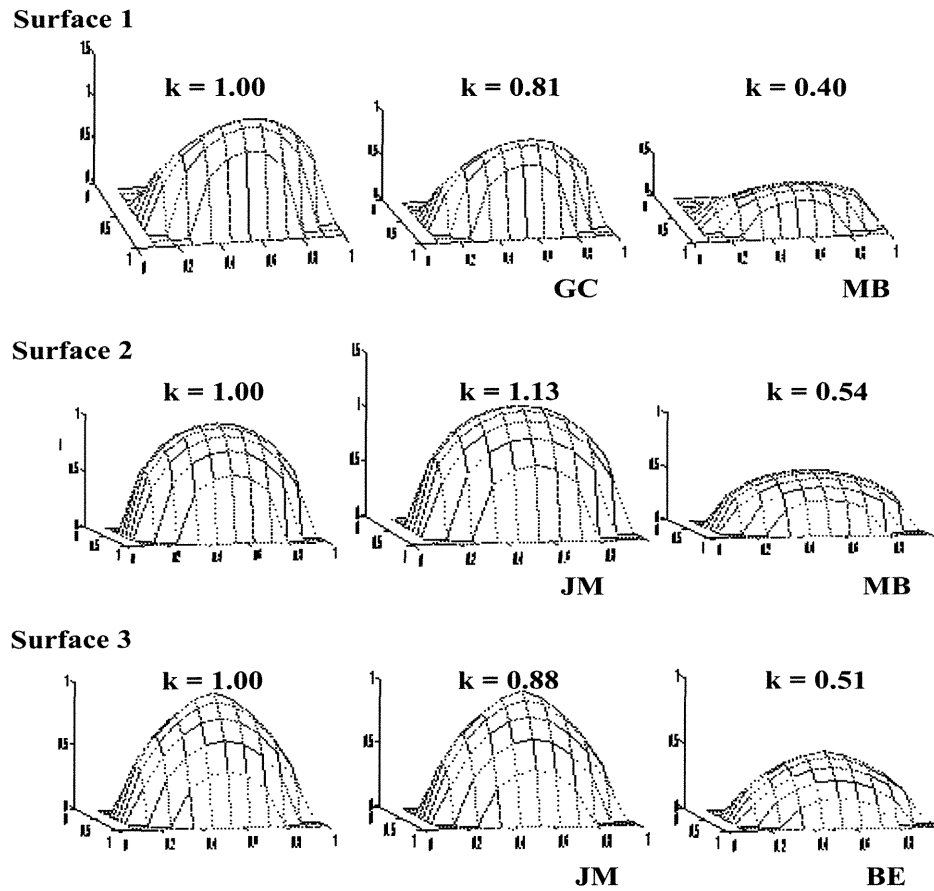


Figure 2.9: Surface plots of the real surfaces are shown in the left column. The middle and right columns show surfaces reconstructed from the data of two subjects for each of Surfaces 1, 2, & 3.

2.4 Tangent Plane Discontinuities

Boundary contours have strong influences on shape perception. We explored the issue of contour influences on shape perception by creating shapes that have tangent plane discontinuities that are not occlusion boundaries (See Figure 2.2, Surfaces 4, 5, & 6). Apparent surface normals were measured using the procedure described in Method 1.

Subjective surface normals are compared with ground truth surface normals in Figure 2.10. Results indicate that subjective perception of nearby surface normals is biased towards the normal to these tangent plane discontinuities, even when the discontinuity projects to a curved line in the image, as in Surface 6.

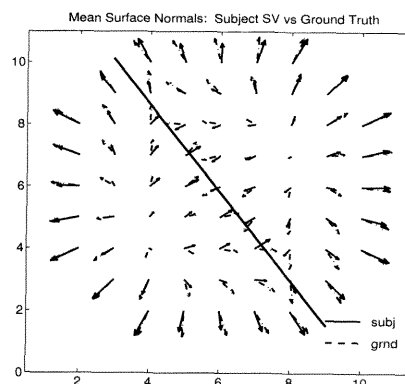
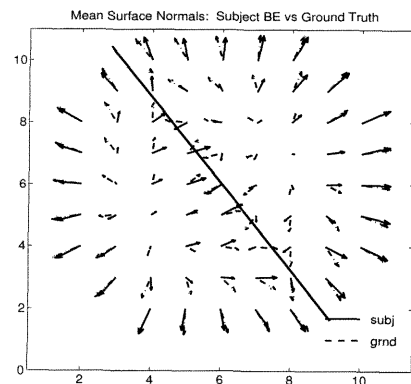
2.5 Scale Effects

Another observation we made from subjective reports was that smaller shapes seemed to be perceived as being more convex, or “shapely,” than larger shapes that are shaded exactly the same way.

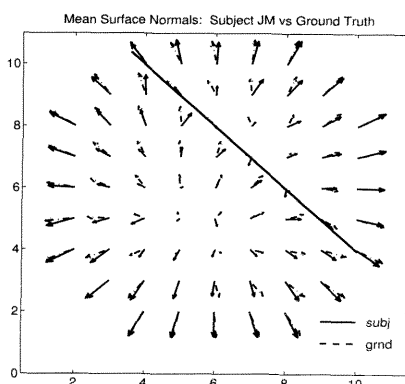
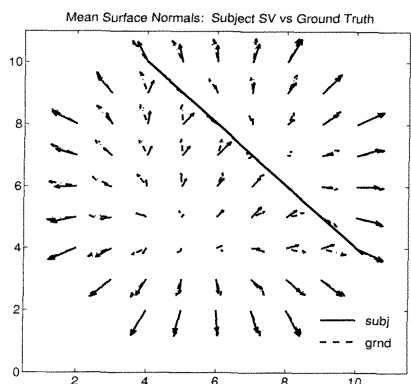
We investigated this phenomenon first using stimuli surfaces 1, 2, & 3 of two sizes, spanning 5 or 10 degrees of visual arc, and estimating the respective depth scalings in the manner described in Section III. These depth scalings are plotted below for different subjects. While there seems to be a trend of reduced depth scalings for the larger stimuli, the finding was not conclusive (See Figure 2.11). We suspected that the effect might be more obvious for even smaller stimuli. Since this gauge measurement method is not amenable for studying shapes much smaller than 5 degrees, we adopted a different paradigm for clarifying this issue (See Method 1).

In the experiment where shaded bubbles were compared with arcs, psychometric curves were generated from the data by tabulating the percentage of time a shaded

Surface 4



Surface 5



Surface 6

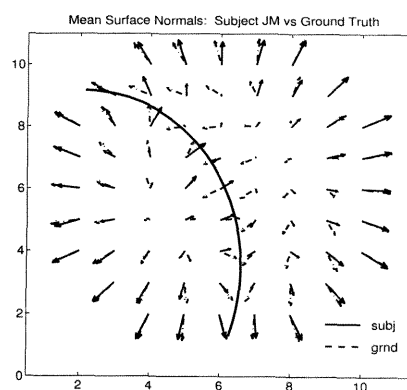
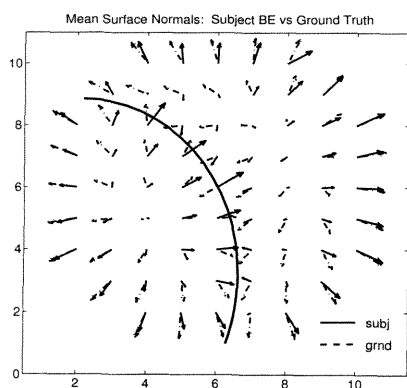
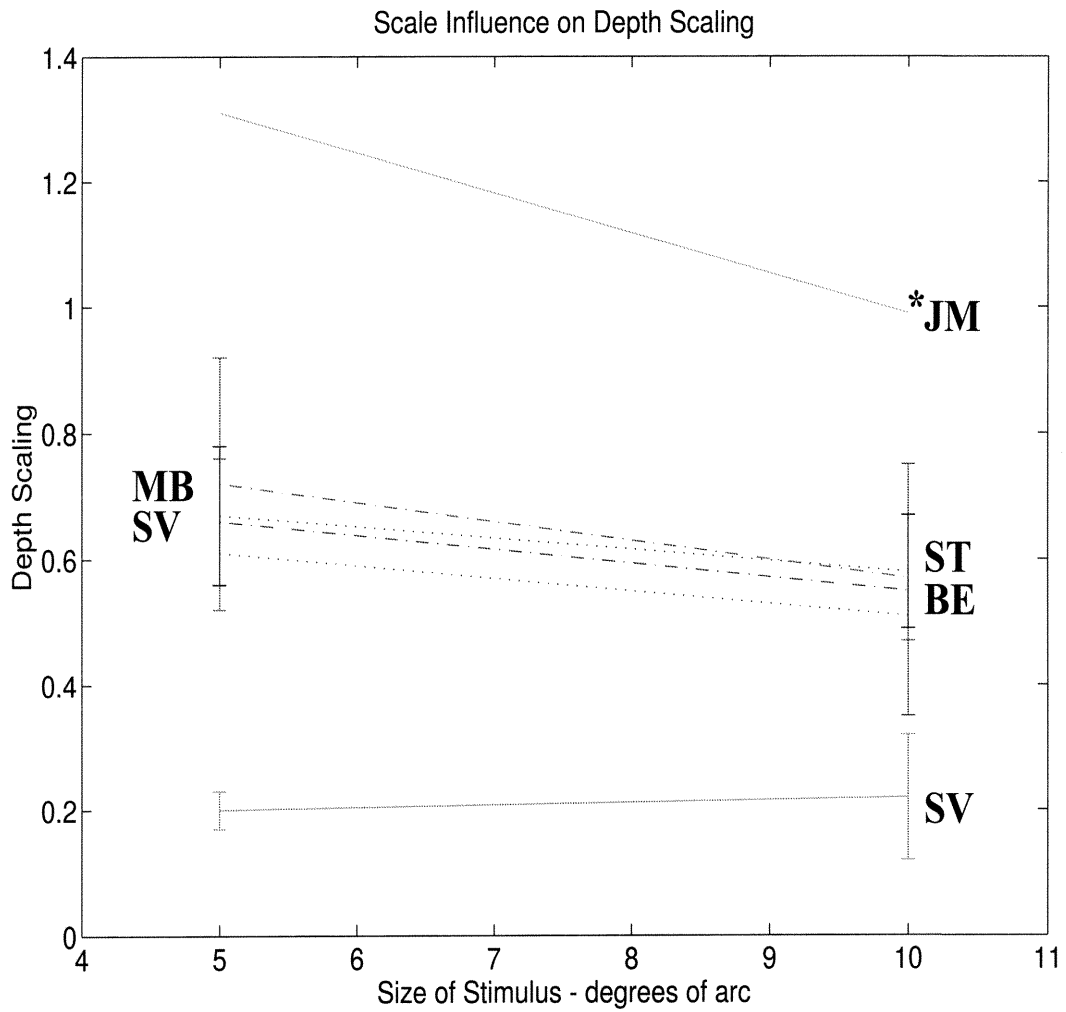


Figure 2.10: Apparent surface normals (solid) for 2 subjects each are shown here for Surfaces 4, 5, & 6 in comparison with the actual surface normals (dotted). The solid line indicates the loci of tangent plane discontinuities for the surface.



*Error bars are unavailable for data sets that included only one sitting for each stimulus size.

Figure 2.11: Influence of image size on perceived depth scaling is shown here for 6 subjects.

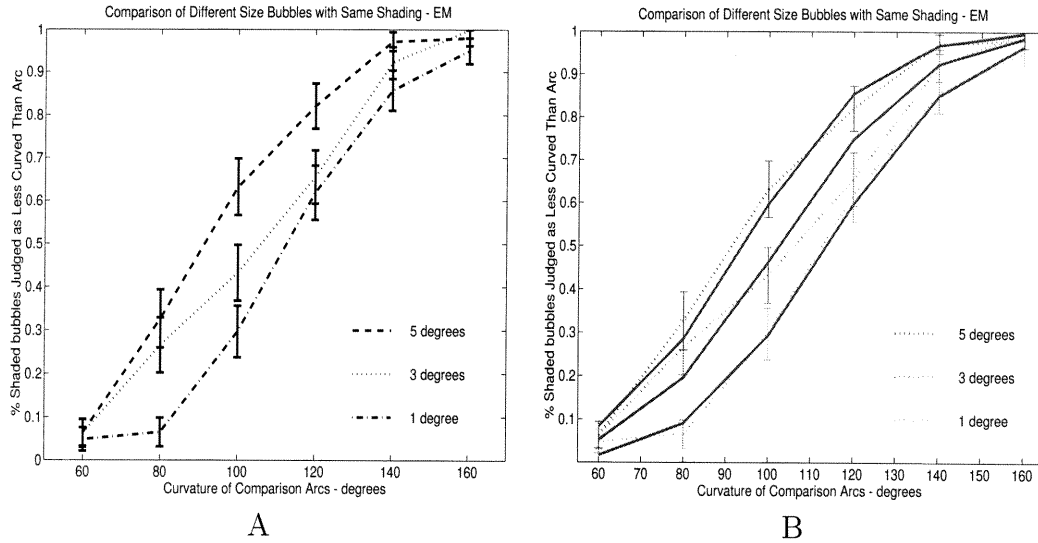


Figure 2.12: Percentage of time a subject judged the bubble to be less shapely than the comparison arcs is plotted in A against the angle subtended by the comparison arcs. B shows the same psychometric curves with fitted error functions superimposed.

bubble of a particular shading and size was judged as being less protruding than the comparison arcs (Figure 2.12A). Error functions are then fitted to the psychometric curves to obtain a 50% index, a fitted mean representing the angle subtended by a line arc that most closely approximates the perceived shape (Figure 2.12B; See Appendix C for more details). These fitted means are plotted against bubble size in Figure 2.13 for 3 subjects, and lines are fitted to these plots. The slope and the variance of the slope, derived by the method given in Appendix C, show that all plots except for Plot D have significant negative slopes.

Figures 2.14 to 2.16 show 3 sets of data where the shapeliness of one shaded bubble was compared with another shaded bubble shown on a subsequent screen. Graphs in the first column depict the results of comparing a 1 degree bubble with another similarly shaded bubble of 1, 3, or 5 degrees in size. The 3 degree bubble was the basis for comparison in the second column, and the 5 degree bubble in the third.

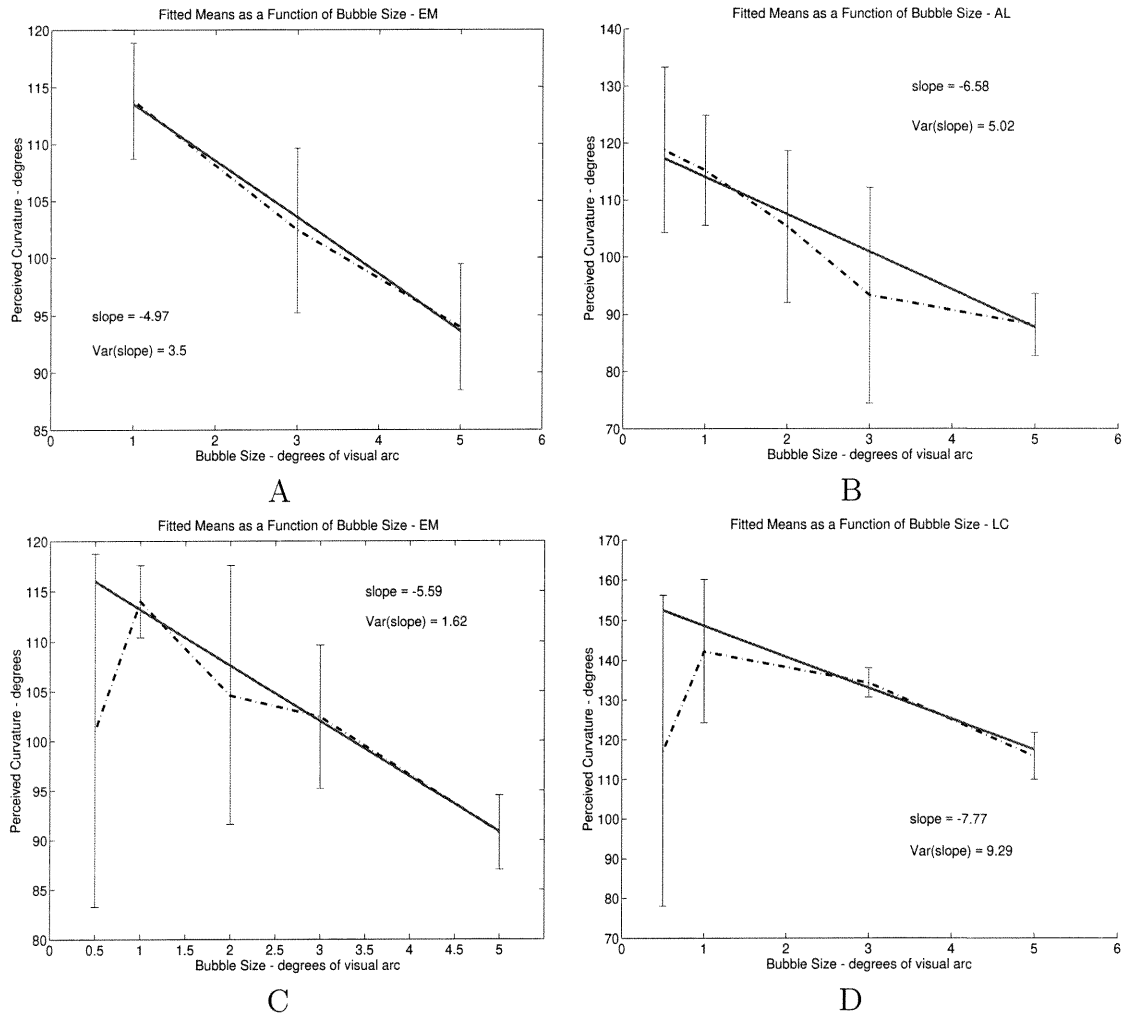


Figure 2.13: The means obtained from fitting error functions to psychometric curves are plotted here with dotted lines against bubble size in degrees of visual arc. Fitted lines are plotted in solid, with the slopes as indicated.

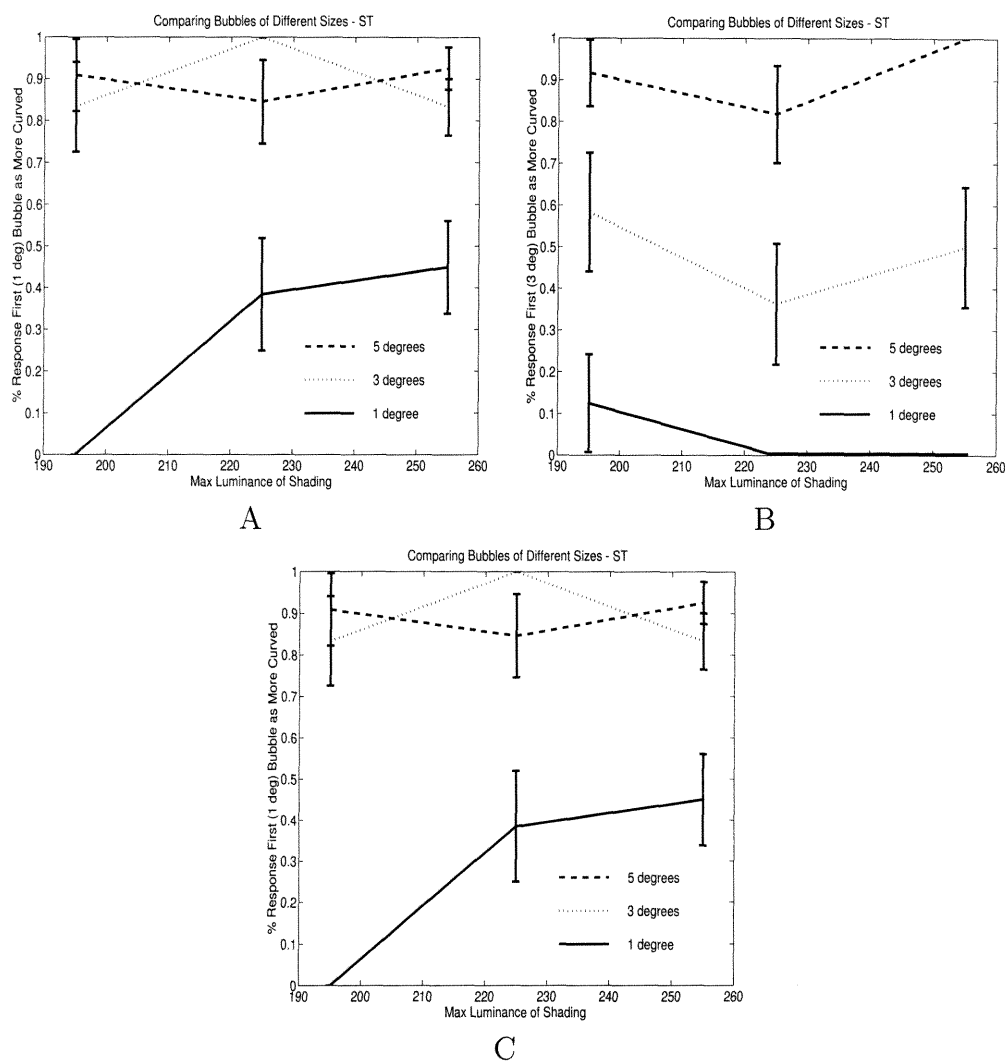


Figure 2.14: Graph A shows the percentage of time that subject ST judged the shaded bubble shown first, a 1-degree bubble, to be more curved than a second shaded bubble, which could be 1, 3, or 5-degrees. In B, the first bubble shown was a 3-degree bubble, and in C, a 5-degree bubble.

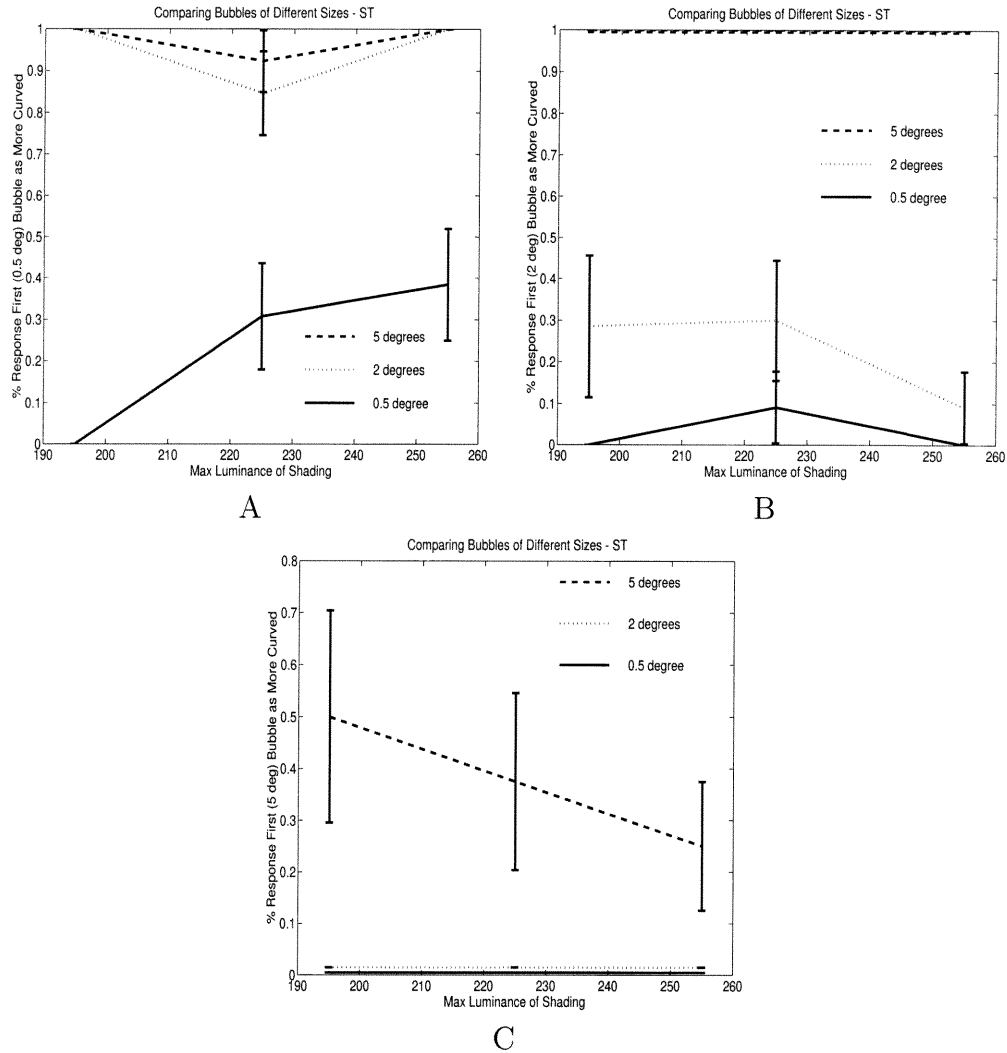


Figure 2.15: Graph A shows the percentage of time that subject ST judged the shaded bubble shown first, a 0.5-degree bubble, to be more curved than a second shaded bubble, which could be 0.5, 2, or 5-degrees. In B, the first bubble shown was a 2-degree bubble, and in C, a 5-degree bubble.

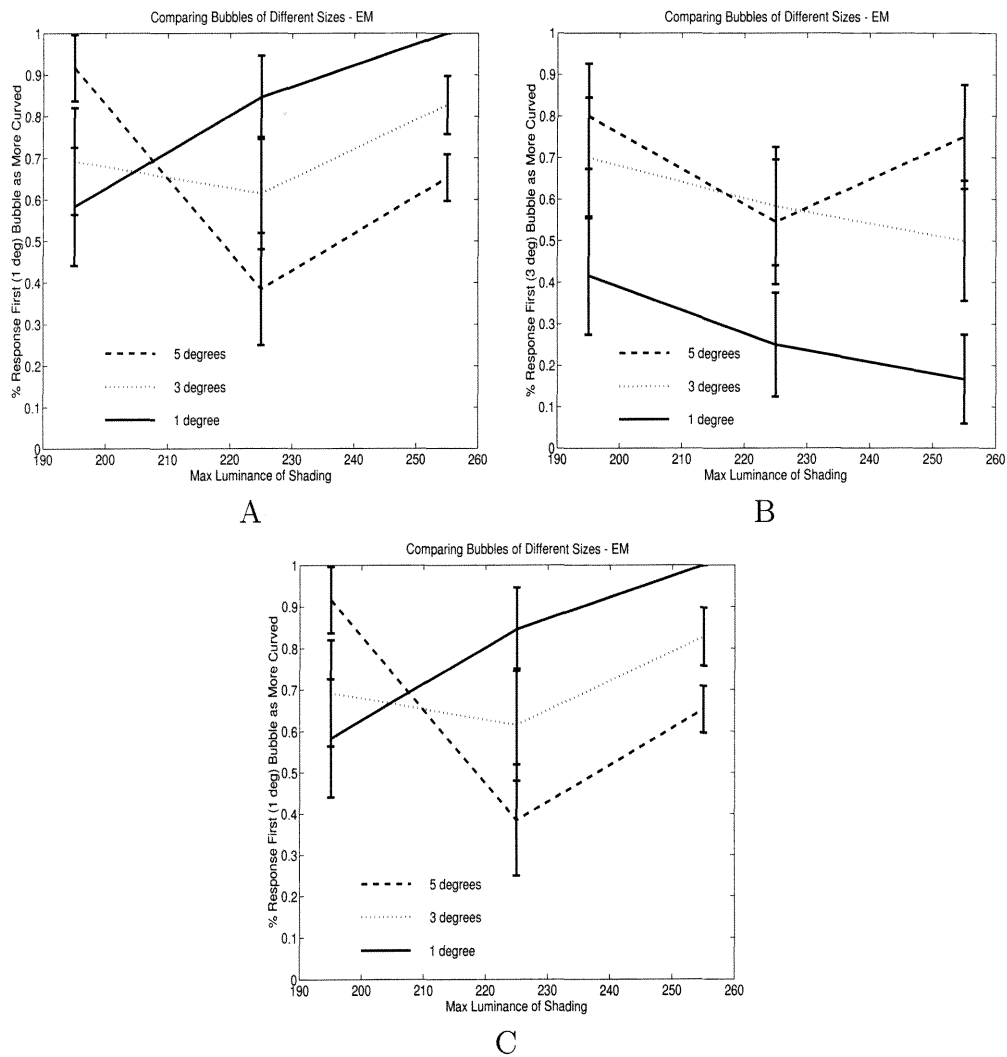


Figure 2.16: Graph A shows the percentage of time that subject EM judged the shaded bubble shown first, a 1-degree bubble, to be more curved than a second shaded bubble, which could be 1, 3, or 5-degrees. In B, the first bubble shown was a 3-degree bubble, and in C, a 5-degree bubble.

The size of the bubbles has a remarkably strong effect on Subject ST. Her data is completely consistent with the notion that smaller shaded shapes are seen as more shapely. For a perfect observer, the line in each graph corresponding to when two identical bubbles are shown should be constant at 0.5. As seen from her left-most graphs, ST has a strong tendency to judge the bubble on the second screen, the one last seen, as more curved when they are actually identical. However, when the first bubble is smaller in size than the second one, the percept of increased shapeliness due to the smaller size overrides even her “recency effect” tendency.

The effect of scaling on perceived shapeliness can be seen clearly with both procedures described in Method 2. Smaller shaded shapes of less than 2 degrees are clearly perceived as being more convex than similarly shaded larger shapes.

2.6 Discussion

In the first part of our experiments, Sections 2.2 & 2.3, we found that subjects display good accuracy in overall shape perception, with an underestimation of depth, as well as good consistency in both tilt and slant directions, with deviations within 5 to 10 degrees. This finding supports the idea that subjects form reliable and stable mental percepts of 3-D surfaces, ones that can persist across both time and varying reflectance functions. This makes sense since we would not want to perceive the same shapes as different on a day-to-day basis, nor would we want our perception of an object’s shape to be dependent upon its material. The fact that this shape constancy holds up under these experimental conditions attests to the validity of our experimental paradigm.

Our finding of good slant consistency is a noteworthy difference from previous

findings (Koenderink *et al.*, 1992). The first possibility that comes to mind is that consistency may be inversely related to surface complexity. However, we found slant deviations to be around 5 degrees for relatively complicated surfaces that have 2 maxima as well as a discontinuous region (See Figures 2.6 & 2.7, Surface 4), as well as for the simpler surfaces.

The small and comparable deviations we found in both the tilt and slant directions show that tilt and slant are useful psychophysical variables for shape perception tasks. They are likely to be closely related to the internal parameters used for the visual encoding of surface orientation (Stevens, 1983a).

Our finding of intra-subject consistency also suggests that the gauge method of quantifying subjective shape perception is a reliable one, and may prove to be a useful technique for future studies of shape perception.

Results from Section 2.4 suggest that there is some robust, low-level process that assumes all tangent-plane discontinuities to be occlusion boundaries, and therefore should have surface normals that are perpendicular to the discontinuity. It is possible that this bias is a local effect that comes about only because the measuring technique itself is very localized, and that there is no corollary deformation of the global percept of the surface. However, we noticed that the discontinuity can in fact cause a change in the global percept. When Surface 6 is viewed in its entirety, there appears to be a dark patch in the center region of the surface, giving the percept of a shallow concavity. If the crease is covered, however, this dark patch promptly disappears. The reader may convince his/herself of this effect simply by obscuring the crease with a finger or a strip of paper. This observation lends support to the idea that the bias we observe in local surface normals does indeed propagate to the global level.

Results from Section 2.5 suggest that there is some optimal stimulus size for

driving the mechanism that underlie shape from shading. The experiments presented here, however, do not establish the lower limit of this optimal size. Some of the data suggests that stimulus size of 0.5 degree does not drive the mechanism as well as that of 1 or 2 degrees, but the error bars are too large for a conclusive statement to be made.

Chapter 3 Processing of Shaded Patterns

The classical studies of preattentive vision have dealt mainly with visual features of the one or two-dimensional world. Typical stimuli included line edges, color, motion, as well as textures and various 2-D shapes. These patterns were sometimes even presented with stereo disparity, but the stimuli were typically neither displayed nor perceived as 3-D shapes (Beck, 1966, 1967, 1982; Olson & Attneave, 1971; Julesz, 1975, 1984; Treisman & Gelade, 1980). More recently, researchers have been turning their attention to the realm of 3-D shapes, and have found, to many's surprise, that preattentive vision does not appear to be constrained to operate only with two-dimensional objects (Ramachandran, 1988; Enns & Rensink, 1990, 1991; Braun, 1990, 1993; He & Nakayama, 1992; Kleffner & Ramachandran, 1992; Sun & Perona, 1996c). There is also recent clinical evidence that shape interpretation occurs very early on in the visual processing hierarchy (Symons *et al.*, 1993). In particular, response-time experiments by Enns & Rensink (1990, 1991) showed that polyhedral targets differing from their distractors in their perceptual 3-D shape “pop-out” with the characteristics of preattentive processing. Moreover, Braun (1990, 1993), using a double-task method, showed that smoothly shaded circular stimuli that resemble spherical bumps or indentations give preattentive pop-out based on 3-D perception.

These results give rise to the following questions:

1. What are the relevant features in these stimuli that allow them to be processed in parallel as 3-D shapes? More specifically, is it the shading itself that is important or is it actually the edge boundaries created by the shaded regions?

2. Is the crucial computation performed locally, e.g. on corner junctions, or is it performed globally upon the entire shape?
3. Is this process a “hard-wired,” local and bottom-up process, or can it be influenced by global and/or contextual information?

The first goal of this paper is to measure 3-D shape pop-out with an experimental paradigm that involves controlled display times and masking. This method will allow us to both verify Enns & Rensink’s results as obtained from response time experiments, and to compare our results with those from previous pop-out and texture segregation experiments involving a similar paradigm (Bergen & Julesz, 1983; Kröse, 1987; Gurnsey & Browse, 1987; Nothdurft, 1991).

The second goal of this paper is to investigate the three questions raised above using two separate sets of experiments. This chapter will deal with the nature and spatial extent of the critical feature. We will be comparing performance on the shaded polyhedral pattern found to elicit pop-out by Enns & Rensink (1990) with that of other patterns differing in specific characteristics of shading, contour, or orientation.

3.1 Methods

3.1.1 Subjects

Five female subjects and five male subjects, all between the ages of 18 and 40, participated in the experiments. Subjects had normal or corrected to normal vision by self-report. All subjects were naive except for one female subject.

3.1.2 Apparatus and Stimuli

Images were generated on a Silicon Graphics Indigo with an 8-bit graphics display and a 16 msec screen refresh rate. Monitor dimensions were 221 mm by 295 mm, with a resolution of 3.47 pixels/mm.

Stimulus screens were viewed binocularly at a distance of 100 cm. Each stimulus screen contained 3, 12, or 24 items of display, with each item spanning approximately 1.5 degrees of visual angle. In screens with 12 and 24 items, spacing between items was approximately 3 degrees, measured from the center of one item to the center of its nearest neighbor, with an additional random jitter of up to 0.3 degrees. For screens of 3 items, the separation was larger, approximately 7.5 degrees, so as to maintain a comparable maximum eccentricity for all display sizes. Grey levels were produced using equal RGB pixel values, ranging from 0 to 256. Stimulus screens had a background RGB value of 80, which gave a measured background luminance of 0.84 cd/m². Shaded stimuli consisted of three regions, each of a different grey level. The corresponding RGB pixel values were 40, 180, and 256. Line stimuli were drawn with RGB values of 0.

3.1.3 Procedure

We used a two-alternative forced-choice stimulus onset asynchrony (SOA) paradigm with masking. Stimulus display times ranged from 16 msec to 400 msec depending on the task, and were followed by a blank inter-stimulus interval (ISI) time of 0, 16, or 26 msec and a 200 msec mask, an example of which is shown in Figure 3.1B. After the mask has disappeared, subjects were asked to report the presence or absence of the target pattern. Within each experiment, the target pattern

was the 180-degree rotation of the distractor pattern. One target was present at random among multiple distractors in 50% of the trials. Target-present trials and target-absent trials had the same total number of patterns. Target position was also randomized, but was constrained to positions of 6.5 degrees of eccentricity or less. Each experimental session consisted of about 1000 trials, presented in blocks of 35 trials, with number of items and duration of display held constant within a block. At the end of each experiment, subjects were asked to describe their perception of the stimuli. Subjects were trained until performance had stabilized, which typically took 2 training sessions.

3.1.4 Data Analysis

Performance for each SOA duration was calculated using d-prime measurements (McNichol, 1972) derived from target-present and target-absent data, resulting in one psychometric curve for each subject in each condition. We make the assumption that, after an initial delay, the variance of the noise affecting the 2AFC decision is inversely proportional to the square root of the duration of the stimulus. Therefore, the psychometric curve was fitted, using a maximum likelihood fitting procedure, to the following model:

$$F(t) = \text{Erf}\left(\sqrt{\frac{t - m}{s}}\right)$$

where m denotes the initial delay, and s is inversely proportional to the steepness of the function. The procedure involved obtaining a two-dimensional likelihood distribution over a range of values for parameters m and s . The pair of values giving the highest likelihood was used to generate a fit to the psychometric curve, and the SOA duration necessary to reach 75% accuracy was calculated from this fitted function.

The spread of the likelihood distribution was used as a measurement of the goodness of fit. The mean SOA across subjects was calculated by averaging the fitted SOA of each subject, weighted by the goodness-of-fit estimate. For cases in which target detection was so difficult that performance did not saturate for even the longest display durations, the fitting was poor. In these instances, we estimated instead, to a 99% confidence level, the minimum SOA duration needed for 75% accuracy. This was done by normalizing the likelihood distribution and finding the σ value up to which the area under the curve equaled 0.01. Minimum SOA durations across subjects were averaged and plotted with an asterisk (*) (See Figure 3.7).

3.2 Experiments

3.2.1 Experiment 1 - Shaded Cubes vs Line Patterns

In this experiment we attempt to establish whether the crucial component for pop-out and, arguably, preattentive shape perception is the oriented edges, the shading, or a combination of the two. We used for this investigation one shaded and two line stimuli previously used by Enns & Rensink in their response time experiments (1990, 1991). The shaded pattern consists of a shaded, Y-junction embedded in a hexagon (See Figure 3.1A). The distractors have an upright Y-junction and are typically interpreted as cubes sitting on a surface with lighting from above. The target has an upside-down Y-junction, and can be seen alternatively as a cube with its bottom side exposed and lit from below, or as a concave corner lit from above. Figure 3.1B shows a typical mask used in the shaded cubes experiments. The line cube and line Y-junction stimuli are shown in Figures 3.1C & 3.1D. Similarly to that for the

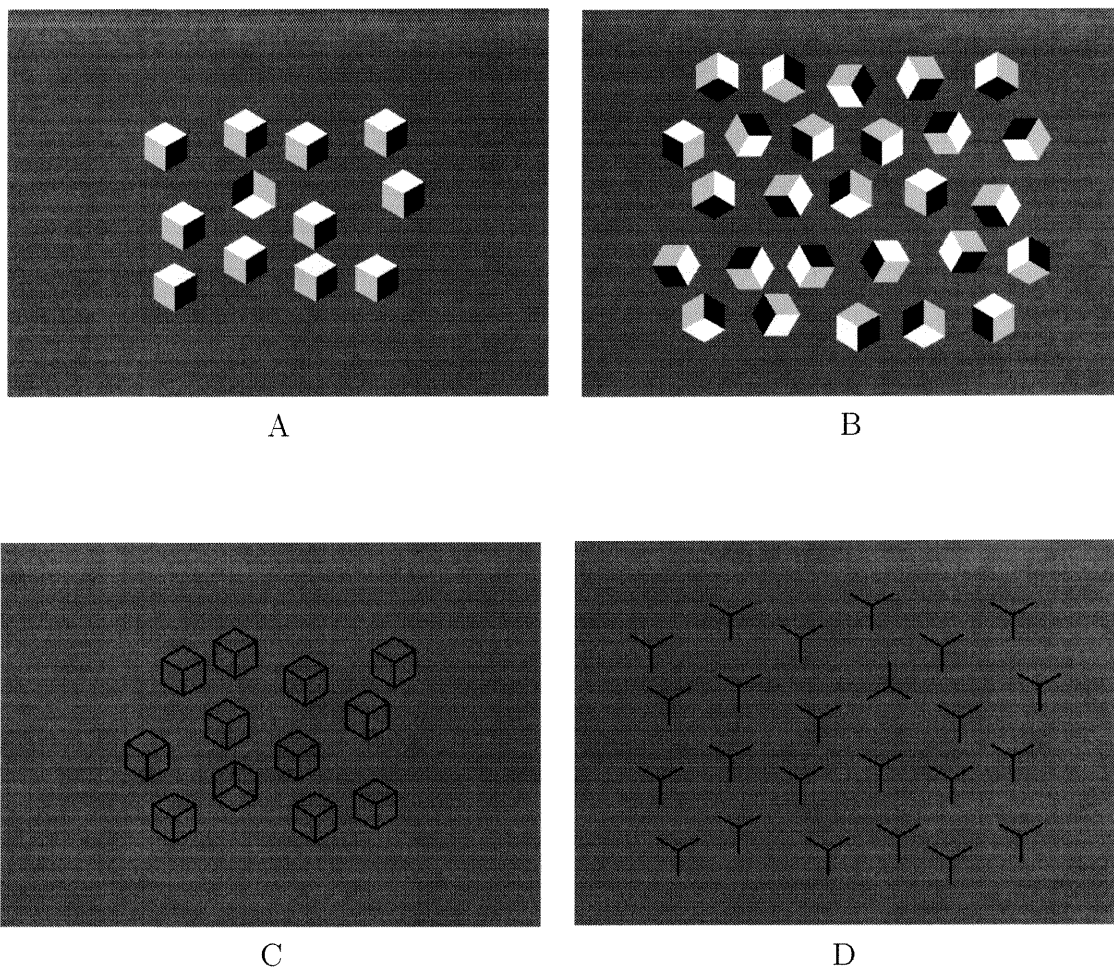


Figure 3.1: Top row shows a sample test screen (A) and a sample mask screen from the shaded cubes experiment (B). Bottom row shows sample test screens from the line cubes experiment (C) and the line Y-junctions experiment (D).

shaded cubes, the masks for these stimuli are composed of 0, 90, 180, and 270-degree rotated versions of the respective distractor pattern.

The average SOA durations necessary for 75% accuracy performance for the shaded cubes, the line cubes, and the line Y-junctions are plotted against the number of display items in Figure 3.2. The average is taken across 7, 5, and 4 subjects respectively. For the shaded cubes, performance is consistently fast across display sizes. The necessary SOA for processing is virtually independent of the number of

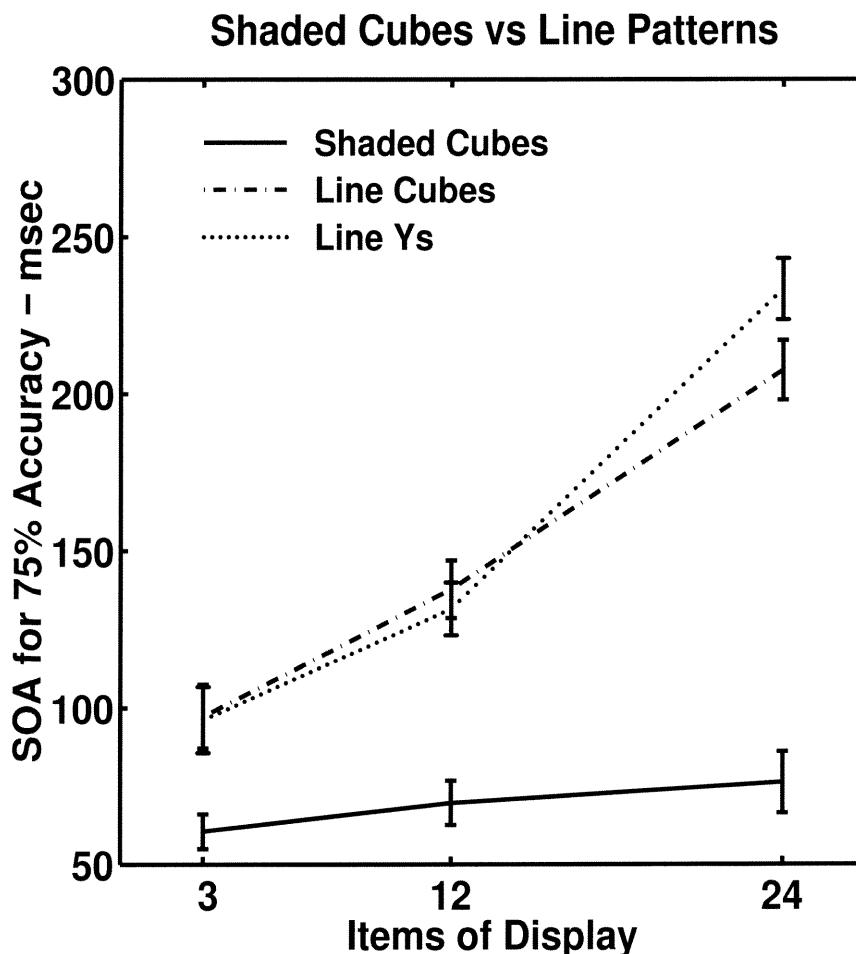


Figure 3.2: SOA for 75% accuracy for 3, 12, and 24 display sizes are shown for each of the three patterns: the shaded cubes, the line cubes, and the line Y-junctions.

distractors. When the data is fitted to a least-squares line, we obtain a slope of 0.8 msec/item, with a standard error of 0.6 msec/item. This result, with $p > 0.09$, is not significantly greater than a slope of zero. In contrast, necessary SOA durations for both the line cubes and the line Y-junctions increase dramatically with the number of distractors presented. The fitted slopes of 5.2 msec/item for the line cubes and 6.0 msec/item for the line Y-junctions are significantly greater than 0, with $p < 0.005$ for both cases.

This behavior, where the fitted slope differs significantly from zero, we will

henceforth refer to as “serial,” as opposed to the relative “parallel” behavior as seen for the shaded cubes. This characterization of visual tasks as parallel or serial, based on the observed independence or dependence of performance on display size, has been made popular by Treisman and collaborators (Treisman & Gelade, 1980; Treisman, 1982; Treisman & Patterson, 1984; Treisman & Gormican, 1988). Not all researchers, however, accept that there is a clean distinction; some prefer the notion that different degrees of task difficulty require different degrees of attention, resulting in a continuum of performance slopes (Duncan & Humphreys, 1989; Wolfe *et al.*, 1992).

3.2.2 Experiment 2 - Shaded 2-D Patterns

The results from Experiment 1 show that shaded cubes, as opposed to line cubes, may be processed fast and in parallel. To investigate whether this fast and parallel processing is actually related to the “3-Dness” of the shaded cube, we used three other patterns that, while shaded in the same black, grey, and white tones as the cubes, do not have typical 3-D interpretations. If these “flat” patterns can also be processed in parallel, then presumably the shaded cubes may also be processed in parallel using 2-D cues only, and the apparent “3-Dness” of the shaded cubes may have nothing whatsoever to do with parallel processing.

Figure 3.3 shows the plots for three such stimuli which we call, respectively, the shaded tiles, the shaded T-junction, and the shaded X-junction, in comparison with the plot for the shaded cubes. The tile and the X-junction patterns were originally used by Enns & Rensink (1990). Data was collected from four different subjects for the tile experiment and five subjects each for the T-junction and the X-junction experiments. For three items of display, performances for the tiles and the T-junctions are, if anything, even better than that for shaded cubes. However, when display size

is increased, necessary SOA durations for all three 2-D patterns are more strongly affected than that of the shaded cubes. The fitted slopes for the tile, T-junction, and X-junction patterns are all greater than 3.0 msec/item, and all significantly greater than 0, with $p < 0.005$. Enns & Rensink's response time experiments (1990) also show that the tile and the X-junction patterns are processed serially.

While the shaded cubes were recognized as convex, lit-from-above cubes without exception, none of the other patterns shown in this experiment prompted 3-D interpretations, except for the shaded tiles, which one subject voluntarily labeled as stairs.

3.2.3 Experiment 3 - Shaded Y-Junctions

The normal orientation shaded cube pattern is composed of a central upright, shaded Y-junction and a hexagonal outline forming three arrow-junctions and three L-junctions. In this experiment, we ask whether this upright, shaded Y-junction is sufficient for fast, parallel 3-D perception, or whether the hexagonal outline that completes the figure of the cube is also necessary. To investigate the effect of the boundary contour, we embedded the Y-junction in three other outlines.

When the Y-junction is embedded in a 30-degree rotation of the outline of the cube, the resulting pattern can be interpreted as a dodecahedron (Figure 3.4, left). Only two of the five subjects recognized it as such, however, while all described it as being less obviously three-dimensional than the shaded cube. Embedding the Y-junction in a diamond-shaped outline results in a pattern that resembles a truncated pyramid (Figure 3.5, left). While all 6 subjects reported seeing the distractors

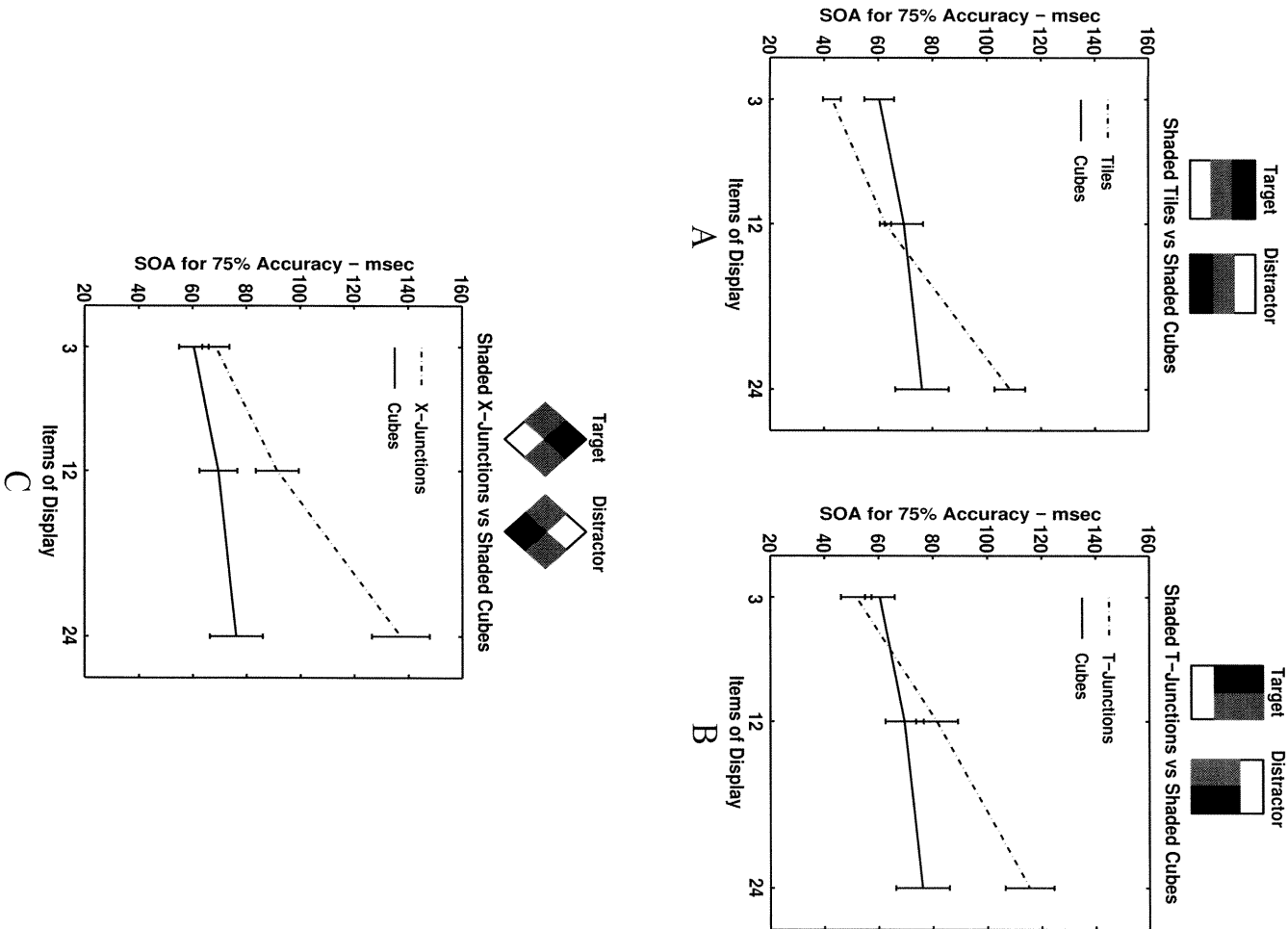


Figure 3.3: SOA for 75% accuracy for three shaded control patterns, shaded tiles (A), shaded X-junctions (B), and shaded T-junctions (C), are shown in comparison with that of the shaded cubes.

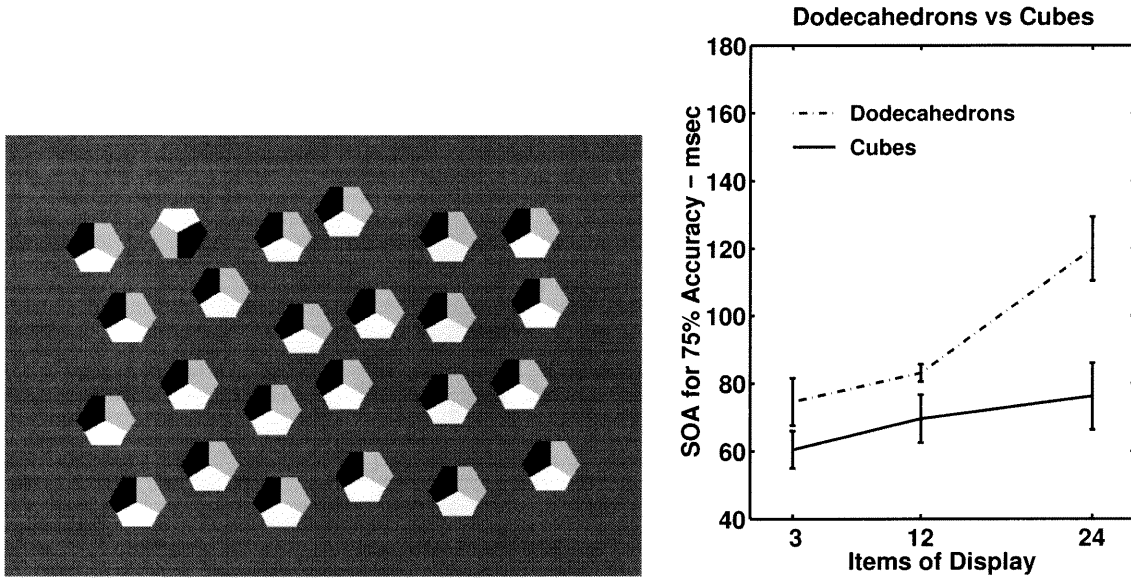


Figure 3.4: The graph on the right summarizes subjects' performance for this hexagonal shaded pattern (left). The performance is parallel for target-absent trials.

as obviously three dimensional at long display durations, they commented that the display was confusing and difficult to organize at short display durations. The results (Figures 3.4 & 3.5, right) show that necessary SOA durations for both these patterns are dependent on display size, with $p < 0.005$, reflecting the perception that these patterns look less consistently 3-D than the shaded cube.

We also embedded the shaded Y-junction in a circular outline (Figure 3.6, left), resulting in a pattern that may be perceived as a 2-D pie chart. In and of itself, this pattern cannot be interpreted as a complete 3-D shape viewed under non-accidental conditions, but, if occlusion is postulated, it can be seen as one corner of a 3-D shape that is being viewed through a circular aperture. Data collected from five subjects revealed that this was the easiest task of all, even easier than the shaded cubes (See Figure 3.6, right). The least-squares-fit line has a slope of 0.6 msec/item, and is not significantly greater than 0 ($p > 0.05$). While this pattern does not actually correspond to any complete 3-D object and may not look convincing 3-D

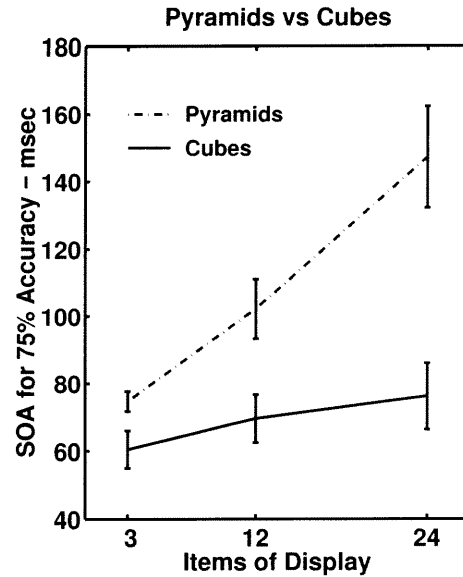
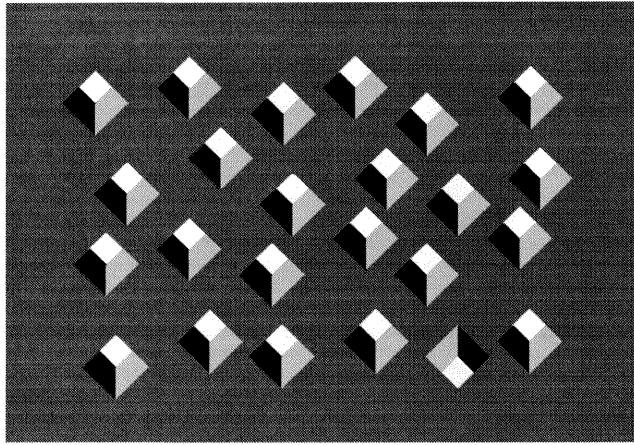


Figure 3.5: Stimulus display screen from the pyramids experiment is shown on the left, and the results are shown on the adjacent graph (right).

when stared at on a static display, subjects reported that, during the experiment, the patterns looked like corners of cubes that are illuminated by circular spotlights or spikes sticking out through a curtain, vividly 3-D in any case.

3.2.4 Experiment 4 - Pop-out Asymmetries

Nine different patterns have been investigated in the previous three experiments. Of these nine, only two, the shaded cubes and the shaded pies, appear to be processed fast and in parallel. It is often the case that when the target and distractor patterns of a parallel task are reversed, performance suddenly becomes serial. Such asymmetries are often used as a diagnostic in the search for primary visual features (Treisman & Gormican, 1988; Williams & Julesz, 1992). In this experiment we investigate whether our two parallel tasks will also become serial when the target and distractor patterns are reversed.

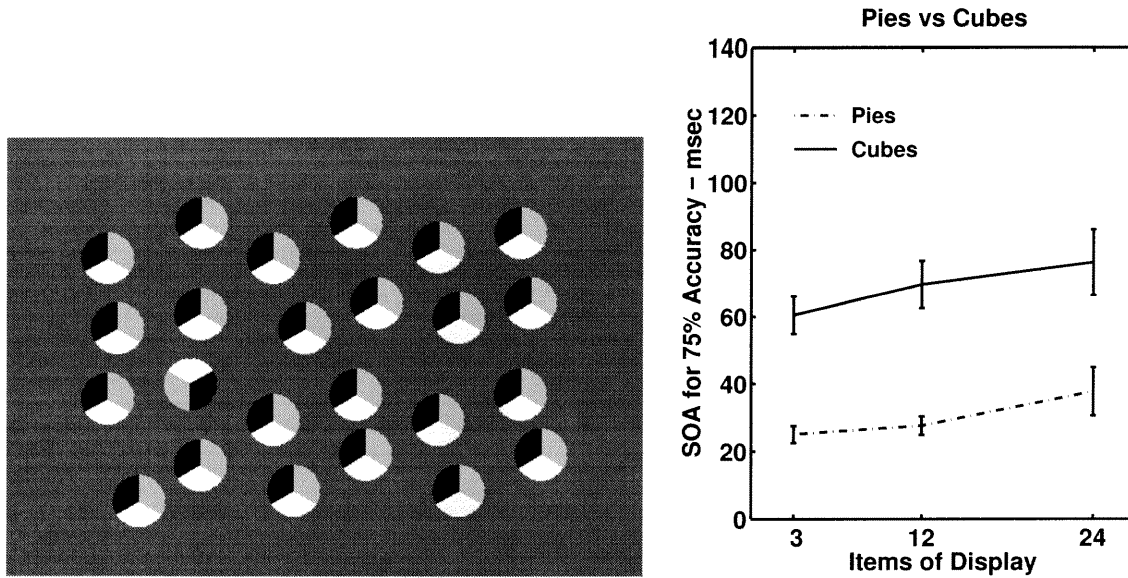


Figure 3.6: A stimulus display screen from the shaded Y-junction in circles experiment (left) and the corresponding performance (right)

Figure 3.7 (left) illustrates the “reverse” condition of the shaded cubes experiment. Compared to the “normal” condition shown in Figure 3.1A, the target and distractor patterns are reversed. Since the target and distractor patterns are related by a 180-degree rotation, the “reverse” condition can be seen simply as a 180-degree rotation of the “normal” condition. We report a dramatic asymmetry in performance between the normal and reverse conditions. For the 24-item case of the reverse condition, performance did not saturate for any of the five subjects even at the longest display duration. The estimated minimum SOA time (See Data Analysis) averaged across subjects is plotted with an asterisk (*) instead. Since reliable 75% correct SOA times cannot be attained from the data, only the data points for the 3- and 12-item displays were used for line-fitting. The resulting fitted line has a slope of 6.5 msec/item, and is significantly greater than 0 ($p < 0.005$).

This asymmetry in performance is correlated with a perceptual asymmetry. While all subjects reported that the distractor in the normal condition experiment looked

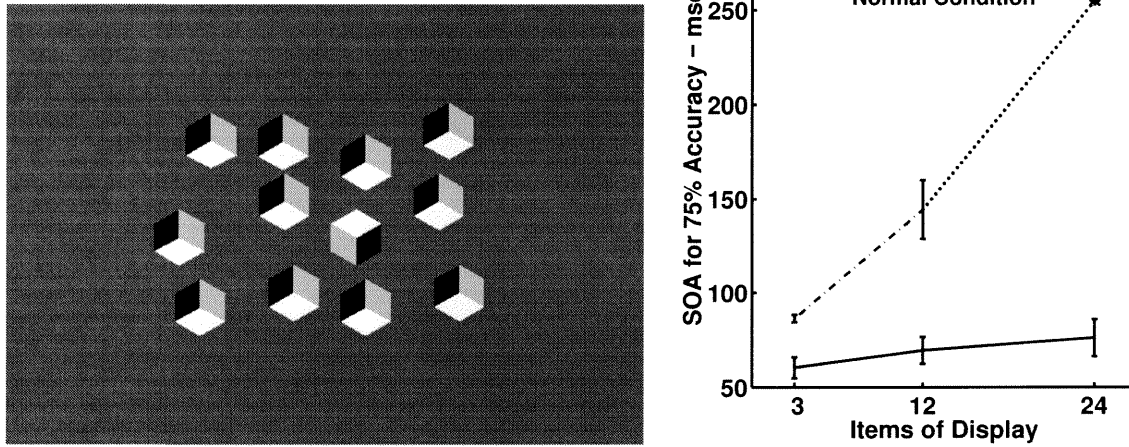


Figure 3.7: “Reverse” orientation shaded cubes are shown on the left. Results from this experiment is shown on the graph at right.

like lit-from-above convex cubes, and were strongly three dimensional, none of the subjects reported 3-D interpretation for the background items in the reverse condition.

Figure 3.8 shows the “reverse” condition of the shaded pies (left) and the resulting performance, averaged across 4 subjects, in comparison with that of the “normal” condition shaded pies (right). The asymmetry is apparent. While performance for the normal-condition shaded pies does not depend significantly upon display size, performance for the reverse-conditions shaded pies does, at a rate of 3.4 msec/item ($p < 0.001$).

3.2.5 Experiment 5 - Y-Junction in Circles

The results so far suggest that the normal orientation shaded Y-junction itself might be a salient feature in 3-D processing. To further explore this idea, we separated the shaded regions of the shaded pie from each other with a gap of about 0.2 degrees. Two gapped patterns were used, one with shaded regions of the same area as the

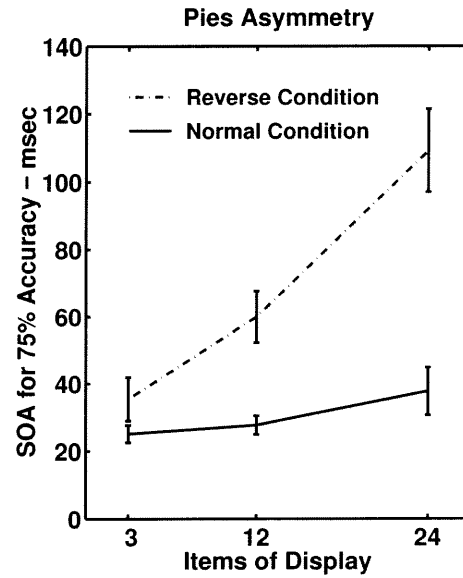
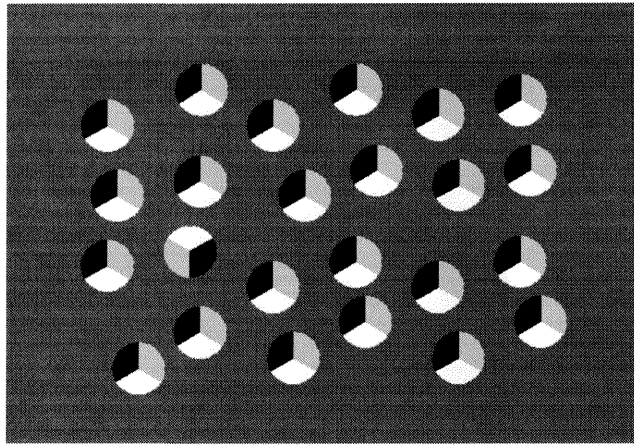


Figure 3.8: “Reverse” orientation shaded pies are shown on the left. Results from this experiment is shown on the graph at right.

no-gap pies (Figure 3.9, top-left), and the other with a total area that was the same as the no-gap pies (Figure 3.9, top-right). Data was collected from 4 subjects for the large gapped pies and 2 subjects for the small gapped pies. Results show that the gaps make a significant difference (Figure 3.9, bottom). Both the large and the small gapped pies have slopes that are significantly greater than 0 ($p < 0.005$). Again, there is a correlation between 3-D interpretation and performance; subjects did not report 3-D interpretation for either sizes of gapped pies.

3.3 Discussion

3.3.1 Parallel Processing of 3-D Shapes

We confirm Enns & Rensink’s finding that the shaded pattern consistent with the interpretation of a top-lit cube can be processed in parallel. We show that this parallel processing is also fast, requiring display durations of less than 80 msec,

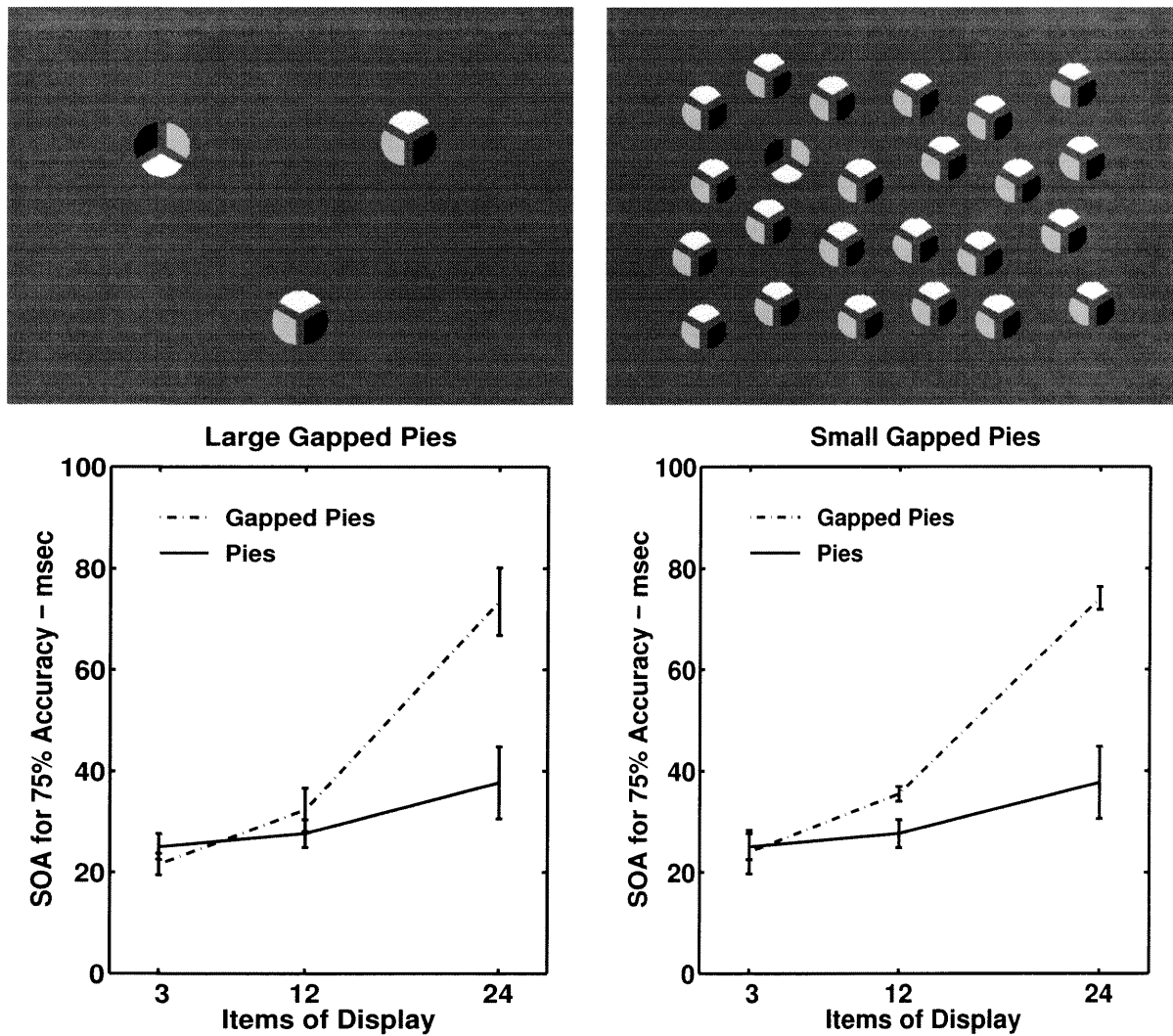


Figure 3.9: Experiments were conducted on gapped Y-junction in circles of two sizes (top). The corresponding performance graph is shown below each pattern.

comparable to fast 2-D pop-out and texture segmentation processes (Bergen & Julesz, 1983; Kröse, 1987; Gurnsey & Browse, 1987; Nothdurft, 1991).

Our results support the idea that mechanisms computing some aspects of 3-D shape are involved in this fast, parallel processing. Both the results of Experiment 2 (Figure 3.3), which involved shaded patterns that do not have 3-D interpretations, and the results of Experiment 4, the asymmetry experiments (Figures 3.7 & 3.8) serve as evidence. Asymmetry in performance is seen for both the shaded cubes

task and the shaded pies task. While the normal orientation tasks are distinctly easy for both, the reverse orientation tasks are significantly more difficult. Since the normal displays and reverse displays are merely 180-degree rotations of each other, they are entirely equivalent in 2-D terms. The vast perceptual difference between the two must therefore lie in their different 3-D interpretations, and a clear perceptual difference was in fact spontaneously reported by all our naive subjects. An analogous asymmetry was found also by Braun (1990, 1993) using smoothly shaded “bubble” stimuli.

3.3.2 Shading Stimuli vs Line Stimuli

In their 1991 paper, Enns & Rensink show that, in a response time paradigm, tasks involving line cubes require search times that increase from approximately 500 msec to 700 msec as the display size is increased from 1 to 6 to 12 items. Other line patterns that do not have 3-D interpretations require search times that increase from 500 to more than 1000 msec. Their conclusion is that the visual system can process line arrow and Y-junctions preattentively, extracting 3-D structure rapidly and in parallel.

Our results, however, indicate otherwise (See Figure 3.2). We find shading to be a crucial cue for driving this fast parallel process. When the shaded cube was replaced by an equivalent line drawing, performance was significantly compromised (Experiment 1) and became serial. We suspect that the difference between our findings is a consequence of our different experimental paradigms. Recent experiments (Sun & Perona, 1996a) suggest that 3-D shape mechanisms driven by line drawings may be used for discrimination when display durations exceed 250 msec. In a response time experiment, where display durations are several hundreds of milliseconds, these

mechanisms may be used to accomplish the task, thereby obscuring the the differences between the shaded and the line-drawing cases.

3.3.3 The Shaded Y-Junction

Results from Experiments 4 and 5 indicate the normal orientation shaded Y-junction to be an important cue for preattentive 3-D processing. We find the shaded pies task to be even easier than the shaded cubes task (Figure 3.6). This extreme ease of processing is disrupted, however, when the contingent shaded regions of the Y-junction are separated by a gap (Figure 3.9).

One interpretation is that the simple and fast 3-D mechanism begins locally by processing the central Y-junctions. If no intrinsic surrounding corner junctions are present, it proceeds quickly to completion. This may apply to the case of the shaded pies. Subjects described this display as resembling convex corners seen through circular apertures. It is possible that the three rounded T-junctions on the surround are perceived as the results of occlusion, and are therefore not considered as intrinsic corners of the figure. In such a case, only the central Y-junction would need to be processed in order to achieve a 3-D percept.

When the surrounding corner junctions fit the configuration of a familiar 3-D shape, as in the case of the cubes, they are integrated with relative ease, at a small cost in processing time (Figure 3.6). When the surrounding corner junctions cannot be easily integrated with the central junction to form a familiar shape, however, this basic mechanism fails. While a 3-D interpretation is possible for both the dodecahedron and the truncated pyramid stimuli, these patterns, containing T-junctions and L-junctions that cannot be readily perceived as resulting from occlusion, are accidental views of their possible physical interpretations (See Figures 3.4& 3.5). The cube

pattern, on the other hand, with a combination of arrow junctions and L-junctions on the surround, is a generic view of the prototypical cube (See Nakayama & Shimojo, 1992). Neither of these two other accidental views are as common and familiar as the generic cube view. This correlates with the subjects' perception that these shapes are unconvincingly 3-D or difficult to interpret during short durations of display.

3.3.4 A Convex-lit-from-above Detector

Asymmetric pop-out is a topic of interest in the study of preattentive vision because it is indicative of the presence of a detector that is specialized for one of the two stimuli, but not both (Treisman & Gelade, 1980; Williams & Julesz, 1992). The asymmetry we find in Experiment 4 suggests that this early vision 3-D mechanism also has a preferred stimulus, either a convex-lit-from-above shape or a concave-lit-from-above shape.

The feature detection theory proposed by Treisman and collaborators argues that if a single detector is used in a pop-out task, the task will be easier when the detector is specialized for the target, rather than the distractors. If the distractors are favored by the detector and the target is not, then the task is predicted to be more difficult. This line of reasoning would explain the asymmetry we observe by postulating the existence of a concave-lit-from-above detector. However, since our subjects consistently reported easy perception of convex shapes when the background stimuli were convex-lit-from-above patterns, and no perception of shape, be it concave or convex, when the background stimuli were concave-lit-from-above patterns, we prefer the hypothesis that the convex-lit-from-above shape computation is the one that is primarily subserved by this early vision 3-D mechanism.

Our preference might be accommodated by an alternative theory: Rubenstein & Sagi (1990) have suggested that asymmetries in pop-out performance have to do as much with the level of “noise” generated by the background as with the “signal” associated to the target. If shading patterns that promote a top-lit, convex percept (such as the normal orientation shaded Y-junction) are preferred by this 3-D mechanism, then, as distractors, they would generate a minimum of background noise. Among such a “quiet” background, the target could be spotted fast and in parallel. On the other hand, if the distractors are shaded patterns that do not promote the preferred interpretation (e.g. the upside-down shaded Y-junction), the background noise level would be high. To detect the signal generated by the target among such a noisy background would then require a serial search. Recent experiments (Sun & Perona, 1996a) suggest that the primary cue for 3-D pop-out is reflectance, rather than 3-D shape or luminance. A percept of top-lit, convex shape leads to discounting of apparent luminance, resulting in a more uniform apparent reflectance. Top-lit, convex distractors would result in a quiet background and easy target detection, while distractors that do not promote such an interpretation would result in a noisy background and difficult target detection.

Chapter 4 Computation of Shape and Reflectance in Early Vision

We previously reported (Sun & Perona, 1996a) that certain shaded patterns which can be interpreted as convex 3-D shapes can be processed at durations as short as 50 msec and promote pop-out, while other similar line or shaded patterns that do not have 3-D interpretation are processed more slowly. We also found that contextual or perspective cues which can enhance 3-D interpretation of the visual scene can also improve performance. These results suggest that there are early mechanisms which subserve the parallel processing of shaded 3-D shapes.

One issue of particular interest is the phenomenon of pop-out asymmetries. While a target consistent with the interpretation of a convex, lit-from-below cube pops out from a background of convex-lit-from-above cube patterns, when the target and distractor patterns are reversed, the task requires serial search (See Figure 4.1; see also Chapter 3, Experiment 2). Such asymmetries have been found with smoothly shaded shapes (Braun, 1990, 1993) as well as with polyhedral shaded shapes (Enns & Rensink, 1990; Sun & Perona, 1996a, 1996c). How might such pop-out asymmetries be explained, and what might they tell us about early vision 3-D mechanisms?

The feature detection theory proposed by Treisman and collaborators argue that if a single detector is used in a pop-out task, the task will be easier when the detector is specialized for the target, rather than the distractors (See Figure 4.2). According to this line of reasoning, there is apparently

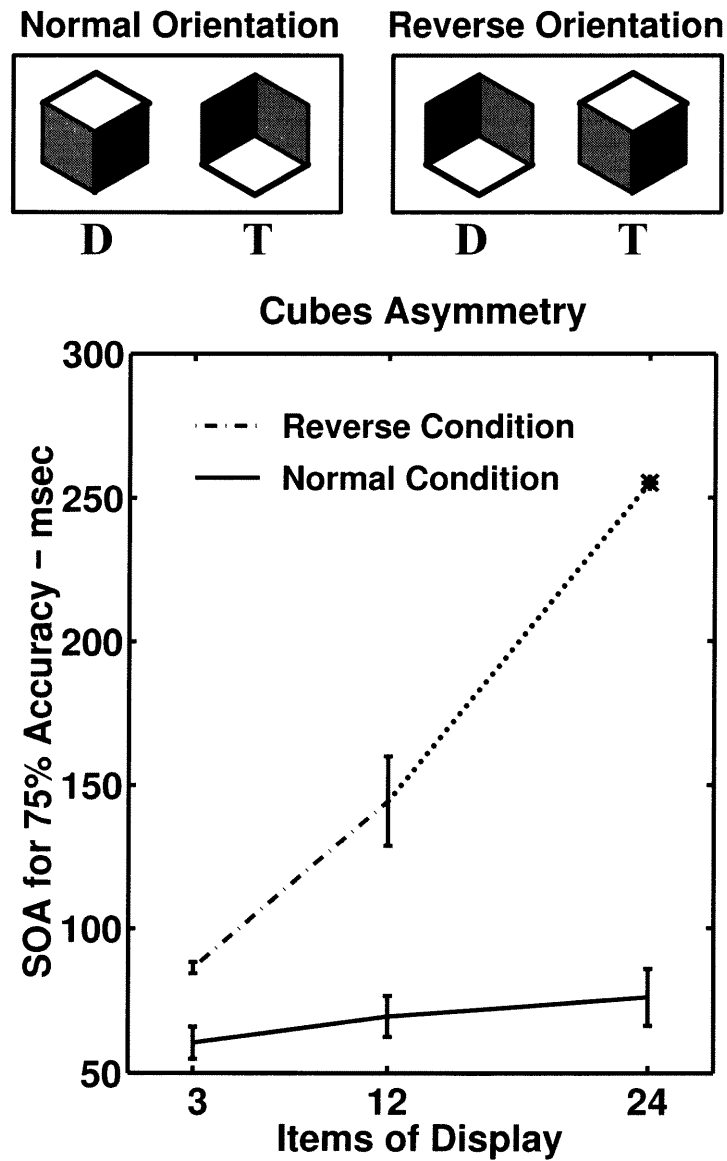


Figure 4.1: Pop-out behavior is seen when the task is to spot an inverted target among upright cubes. When the target and distractor patterns are reversed, serial search behavior is seen.

Pop-out Asymmetry

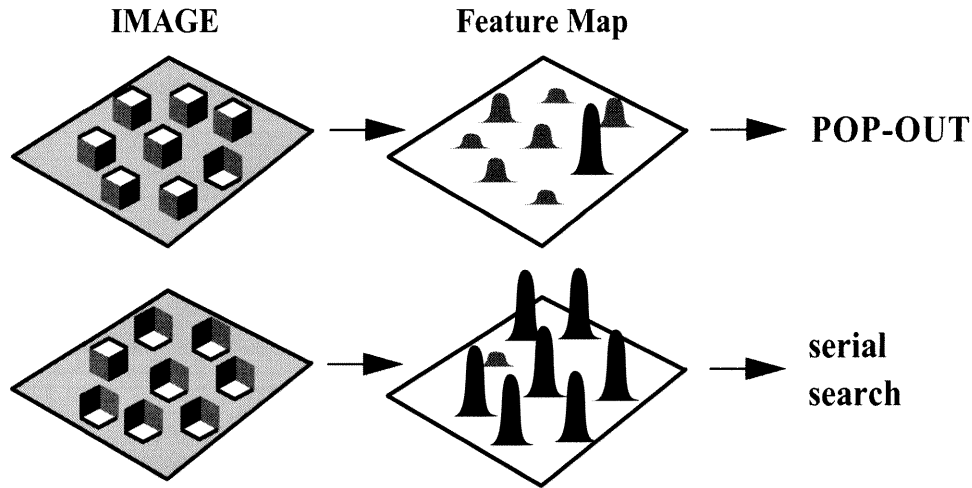


Figure 4.2: According to the Feature Detection Theory proposed by Treisman & collaborators, pop-out occurs when the target, but not the distractors, is recognized by the feature detector.

1. a detector that is “tuned” to the lit-from-below stimulus rather than the lit-from-above one. *However,*
2. our subjects report easy perception of convex shapes when the background stimuli were convex-lit-from-above and no perception of shape be it convex or concave, when the background stimuli were convex-lit-from-below. *Yet,*
3. research shows that pop-out is associated with 3-D perception.

So how do we solve this conundrum?

We suggest, as an alternative, that fast discrimination is based mainly on apparent reflectance. Perceptions of shape and reflectance are closely related (Figure 4.3). When a pattern is seen as flat, its luminance pattern is perceived as due to changes in reflectance. When a pattern is seen as having 3-D shape, however, the changes in its luminance pattern are discounted to some extent as shading due to shape, and its reflectance may be perceived as relatively constant. If apparent reflectance,

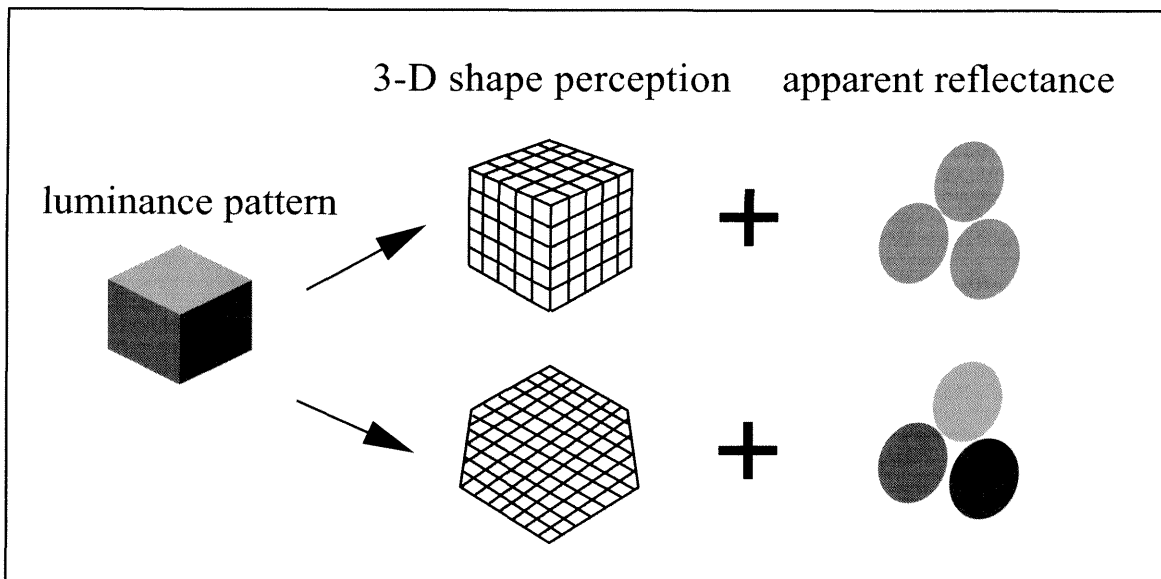


Figure 4.3: The perception of shaded patterns may be represented with a shape map and a reflectance map. If the shaded pattern leads to the perception of a 3-D shape, then the apparent differences in luminance are perceived as shading, and the resulting apparent reflectance is uniform. If the perception is of a flat surface, then the apparent luminance differences are attributed to changes in reflectance.

rather than shape or brightness, is used as the feature for discrimination, a flat, multiple-reflectance target pattern in a background composed of single-reflectance 3-D shapes would be discriminated with ease (Normal Orientation Condition, Figure 4.1). Conversely, if the distractors were perceived as flat, the background would be perceived as having multiple reflectances. Among such a background, the single-reflectance target would be difficult to spot (Reverse Orientation Condition, Figure 4.1).

Our hypothesis: Apparent reflectance is computed in parallel with 3-D shape, and is the primary cue for fast pop-out.

4.1 Methods

Stereo images were generated on a Silicon Graphics Indigo² with stereo capabilities and viewed using SGI *CrystalEyes* goggles. Stimulus screens were viewed at a distance of 20 inches. The vertical height of the screen subtended 30 degrees of visual angle. Each stimulus pattern spanned approximately 1.5 degrees. Items were displayed on randomized depth planes so that the target could not be discriminated based solely on disparity information.

We used a 2afc stimulus onset asynchrony (SOA) paradigm with masking. Each trial began with a cueing screen that contained patterns displayed at the same depth planes as the ones in the test screen that followed. This allowed the subjects to fuse the test screen and achieve stereopsis rapidly. For Experiments 1 & 2, the test screen contained 24 items with one target present among multiple distractors in 50% of the trials. Test screens in Experiment 3 contained either 12 or 24 items. Display duration of the test screen ranged from 80 to 250 msec, followed by a blank ISI of 30 msec and a mask of 200 msec. At the conclusion of the mask, subjects were asked to report the presence or absence of the target pattern. Experiments were presented in blocks of 20 to 35 trials, with the target-distractor pair and the display duration held constant within a block. Each session contained about 1000 trials, lasting approximately one hour.

Data was collected from 5 naive subjects Experiment 1 & 2, and 3 naive subjects in Experiment 3. For every task, target-present and target-absent performances, represented by percentages correct, were used to obtain a measurement of d' for each subject. An overall percentage score is calculated from this d' measurement, and averaged across subjects to obtain a mean percentage score for the each task in

Experiment 1 & 2 (See McNichol, 1972). For Experiment 3, the display duration was varied systematically in order to determine the time necessary to reach 75% correct performance (data analysis as in Chapter 3).

4.2 Experiments

4.2.1 Experiment 1 - Shaded Triptych Patterns

If our hypothesis is correct, 3-D shape could be used to decrease the saliency of a target. In Experiment 1, we used triptych patterns that are easy to discriminate when perceived as flat. 3-D shape was imposed onto both the distractor and the target patterns (Conditions 1 & 2, Figure 4.4A) using stereo disparity cues and perspective projection. To control for the effects of spurious factors not relevant to our hypothesis, we constructed Conditions 3 & 4 (Figure 4.4A). Condition 3 controls for the decrease in uniformity caused by the perspective projection. Condition 4 controls for the possibility that the mere addition of 3-D shape hampers performance by increasing noise and complexity of the task. All conditions were shown at a display duration of 120 msec.

Prediction

Our hypothesis predicts that, when an appropriate 3-D shape is imposed, the differences in luminance of the vertical bands will be discounted, to some extent, as shading due to shape. Instead of flat patterns with distinctly different luminance patterns, the target and distractor patterns will be perceived as 3-D shapes of comparable reflectances. Hence, they will be less distinguishable from each other. If our hypothesis is correct, the task involving 3-D stimuli should be *more difficult* than the

task involving flat ones.

Results

As predicted, we find that the addition of 3-D shape compromises discrimination significantly. Figure 4.4B shows the combined results of 5 subjects performing at a display duration of 120 msec. The 3-D task (1) is significantly more difficult than the flat one (2). We find that neither control conditions (3 & 4) is any more difficult than Condition 1, indicating that the impairment in performance is not due to the decrease of display uniformity or the increase of noise and/or complexity. Impairment is seen, as in Condition 1, only when the luminance pattern is consistent with the 3-D shape, and may therefore be perceived as shading, rather than as changes in reflectance. The results of Experiment 1 support our hypothesis.

4.2.2 Experiment 2 - Shaded Corner Patterns

If shape contributes to pop-out by discounting perceived luminance differences, then it should be possible to influence performance both positively and negatively by manipulating the 3-D shapes of the target and distractors separately. In Experiment 1, shape was used to decrease the saliency of the target, leading to impairment of performance. In Experiment 2, we attempted to increase the saliency of the target through the use of 3-D shape. Displays in this experiment were composed of simple patterns of two grey levels that may be perceived as shaded corners, shown at a display duration of 180 msec. This 3-D shape was imposed onto either the distractors (Condition 1, Figure 4.5A) or the target (Condition 2, Figure 4.5B) using stereo disparity cues and perspective projection. We constructed a third condition in which both the target and the distractor patterns were flat. The dark and light bands of

Experiment 1

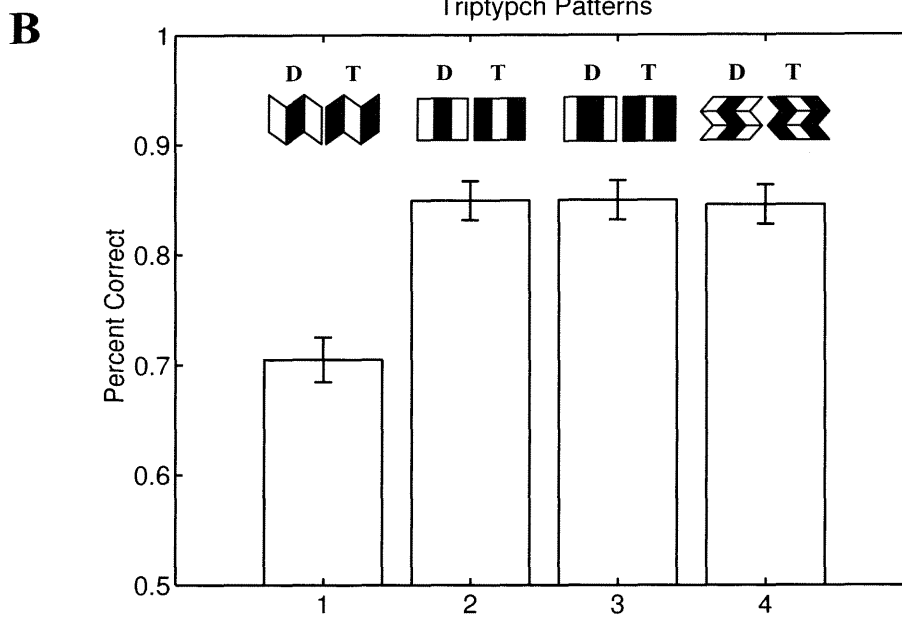
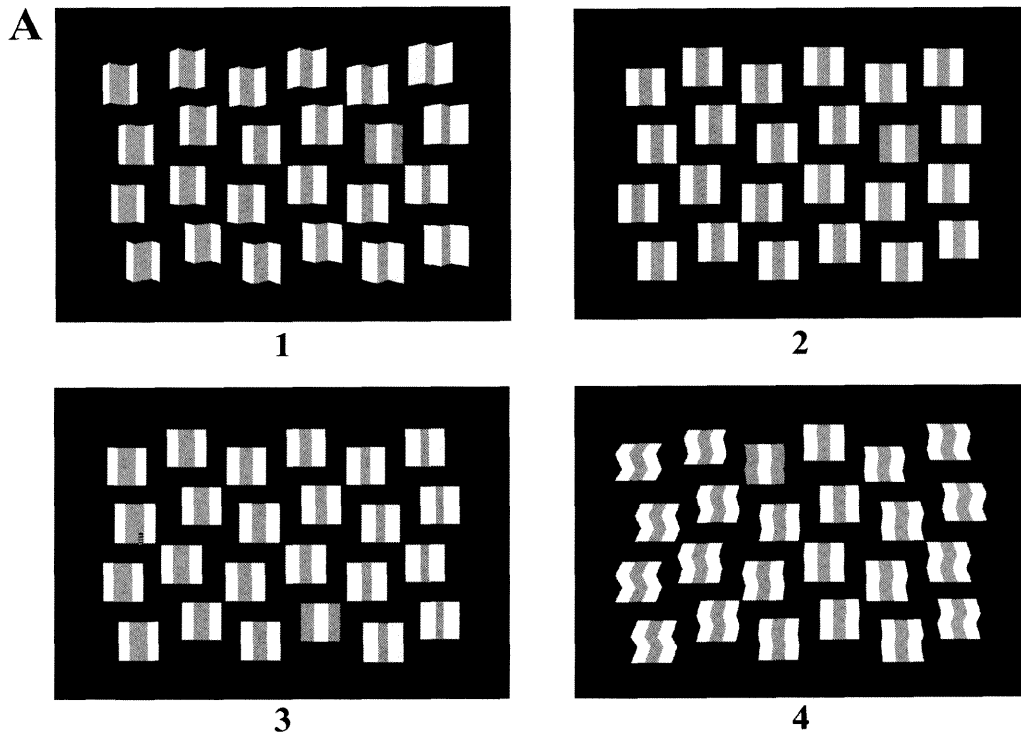


Figure 4.4: The right eye's view of Conditions 1-4 in Experiment 1 is shown in A. B shows data collected from 5 naive subjects performing at a display duration of 120 msec. Significant impairment of performance is seen only when the 3-D shape was perceived as single-reflectance, and the luminance pattern can be interpreted as shading (Condition 1).

the distractors in this condition have widths that are dependent upon the location of the distractor, in order to mimic the perspective effects seen in Condition 1. Test screens were displayed for 180 msec, followed by masking. To isolate the effects of shape from shading, we constructed also the wire-frame equivalents of Conditions 1 & 2 (Conditions 4 & 5, not shown), displayed at a duration of 250 msec.

Prediction

Our hypothesis predicts Condition 1 to be easier than Condition 2. A flat target with an “undiscounted” luminance pattern among distractors whose luminance pattern are somewhat discounted by shape should be relatively easy to spot. On the other hand, a single-reflectance 3-D target in a background noisy with luminance differences should be more difficult to spot.

Results

We find that, for the shaded patterns, there is an asymmetry in performance between Conditions 1 & 2 (Figure 4.5C); it is easier to spot a flat target among 3-D distractors than to pick out a 3-D target among flat distractors. If discrimination were based directly on 3-D shape, the same asymmetry should occur for the wire-frame stimuli. However, we see from Conditions 3 & 4 (Figure 4.5D) that the asymmetry is in the *opposite* direction. In this case, it is easier to pick out a 3-D target among flat distractors than vice versa. Moreover, for the wire-frame stimuli, discrimination is quite slow by comparison; display durations were 180 msec for the shaded conditions and 250 msec for the wire-frame ones. These findings are very difficult to reconcile with the view that fast discrimination is based directly on shape, and are instead consistent with our hypothesis that reflectance differences are important in determining

Experiment 2

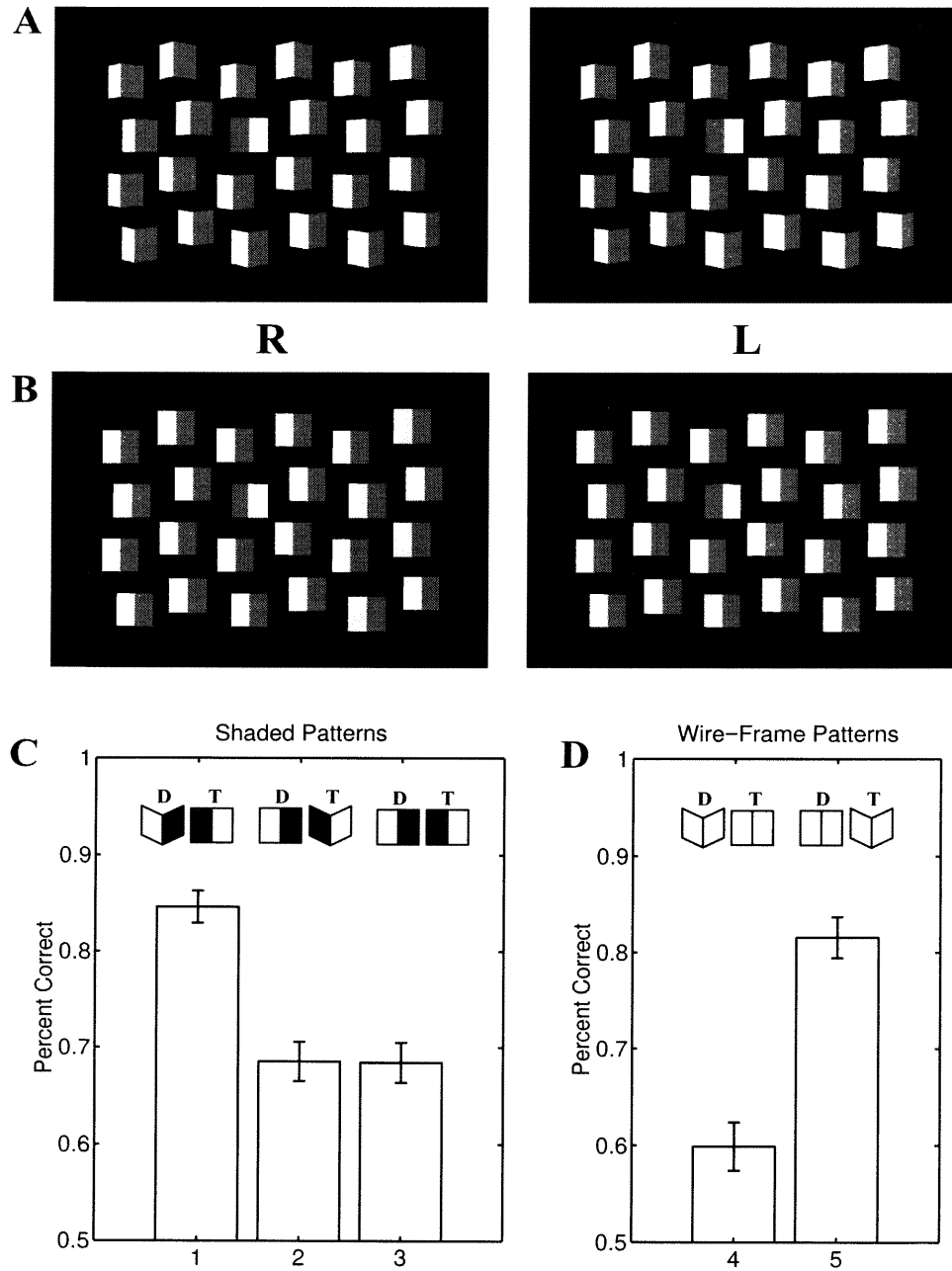


Figure 4.5: Cross-fusing stereo pairs for Condition 1 (A) and Condition 2 (B) of Experiment 1 are shown. (C) Performance averaged over 5 naive subjects for Conditions 1, 2, & 3 (both distractors and target were flat) indicate a marked asymmetry between Conditions 1 & 2. (D) Wire-frame patterns were used as controls. The direction of asymmetry is opposite to that of the shaded conditions. Display duration was 250 msec.

pop-out.

4.2.3 Experiment 3 - Parallel or Serial?

In Experiment 3, we investigated whether the large performance differences seen in Experiments 1 & 2 are correlated with a qualitative change from serial to parallel processing (Triesman & Gelade, 1980). In serial visual search tasks, the time necessary for target discrimination increases significantly when the number of distractors is increased. In parallel tasks, on the other hand, the necessary time is not significantly dependent on the number of distractors.

We tested the two key stimulus pairs from each of Experiments 1 & 2 on 3 naive subjects with displays containing either 12 or 24 stimulus patterns, using the same paradigm as in Experiments 1 & 2. The display duration was varied systematically in order to determine the time necessary to reach 75% correct performance (data analysis as in Chapter 3).

Results

For the triptych patterns from Experiment 1 (Figure 4.6A), the flat patterns elicited parallel performance while the 3-D patterns did not. For the stimuli from Experiment 2 we found that when the target, but not the distractors, had 3-D shape, performance was serial (Figure 4.6B). When the distractors, but not the target, had 3-D shape, performance approached parallel behavior. Overall, these results suggest a qualitative change in processing behavior that is correlated with that of Experiments 1 & 2.

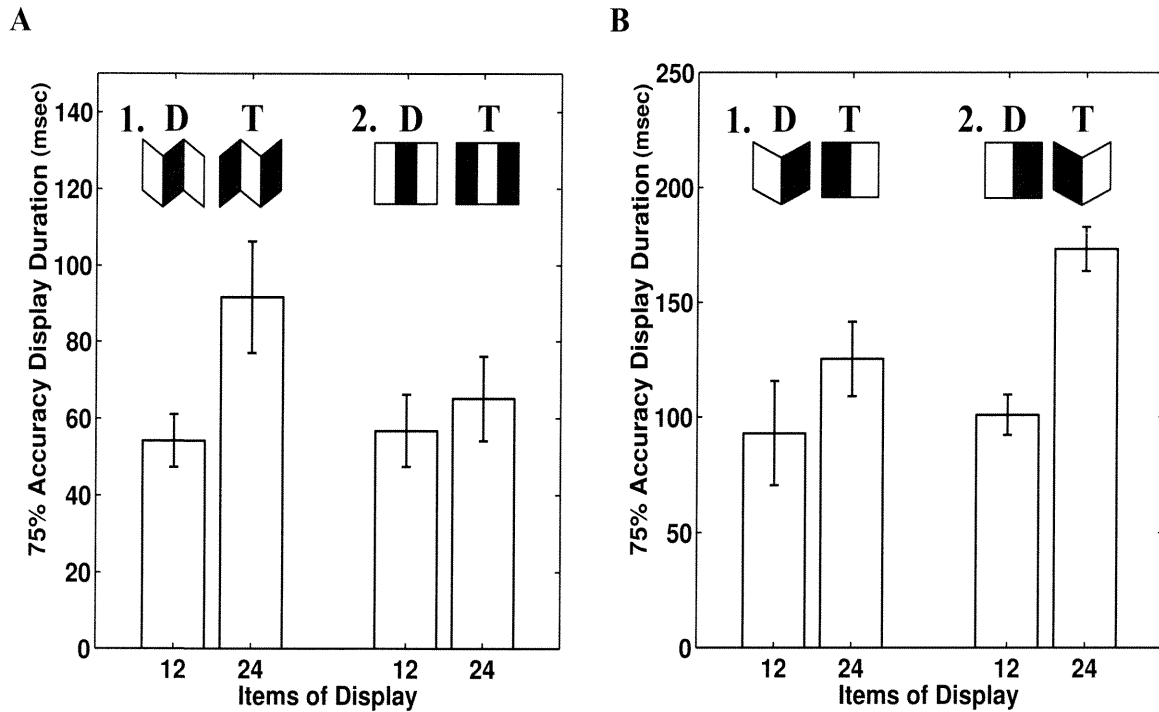


Figure 4.6: (A) Stimulus patterns from Conditions 1 & 2 of Experiment 1 were tested in two display sizes. When the display size was increased from 12 to 24, the 75% accuracy display duration increased significantly for Condition 1 but not for Condition 2. (B) For Conditions 1 & 2 in Experiment 2, when the display size was increased from 12 to 24 items, there was a significant increase in the 75% accuracy display duration for Condition 2 but not for Condition 1.

4.3 Discussion

Our hypothesis:

Early vision 3-D mechanisms subserve fast reflectance calculations, participating in pop-out by discounting luminance differences.

The evidence from our experiments supports the hypothesis:

1. Experiment 1 shows that 3-D shape perception from stereo cues of both the target and the distractors led to impairment of performance. Control experiments indicate the most likely explanation to be the discounting of luminance differences.
2. Experiment 2 shows that by conferring 3-D shape only to distractors or only to targets result in opposite effects. When the distractors are 3-D, luminance differences are discounted to an extent, rendering the flat target easy to spot. When the distractors are flat and the target is 3-D, however, the noisy background together with an inconspicuous target make a difficult task.
3. Experiment 3 shows that the improvements and impairments observed in Experiments 1 & 2 are correlated with qualitative changes in processing behavior, from parallel to serial and vice versa.

The discounting of luminance difference when a shaded 2-D pattern is interpreted as a 3-D shape can be seen even during stable viewing conditions. In Figure 4.7, a “C” composed of upside-down cubes is embedded within upright cubes. The luminance of the darkest side of the upside-down cubes is varied systematically from light to dark. The task is to choose the one that has a darkest side that is identical to that of the upright cubes. While the correct answer is actually the fifth upside-down cube in the

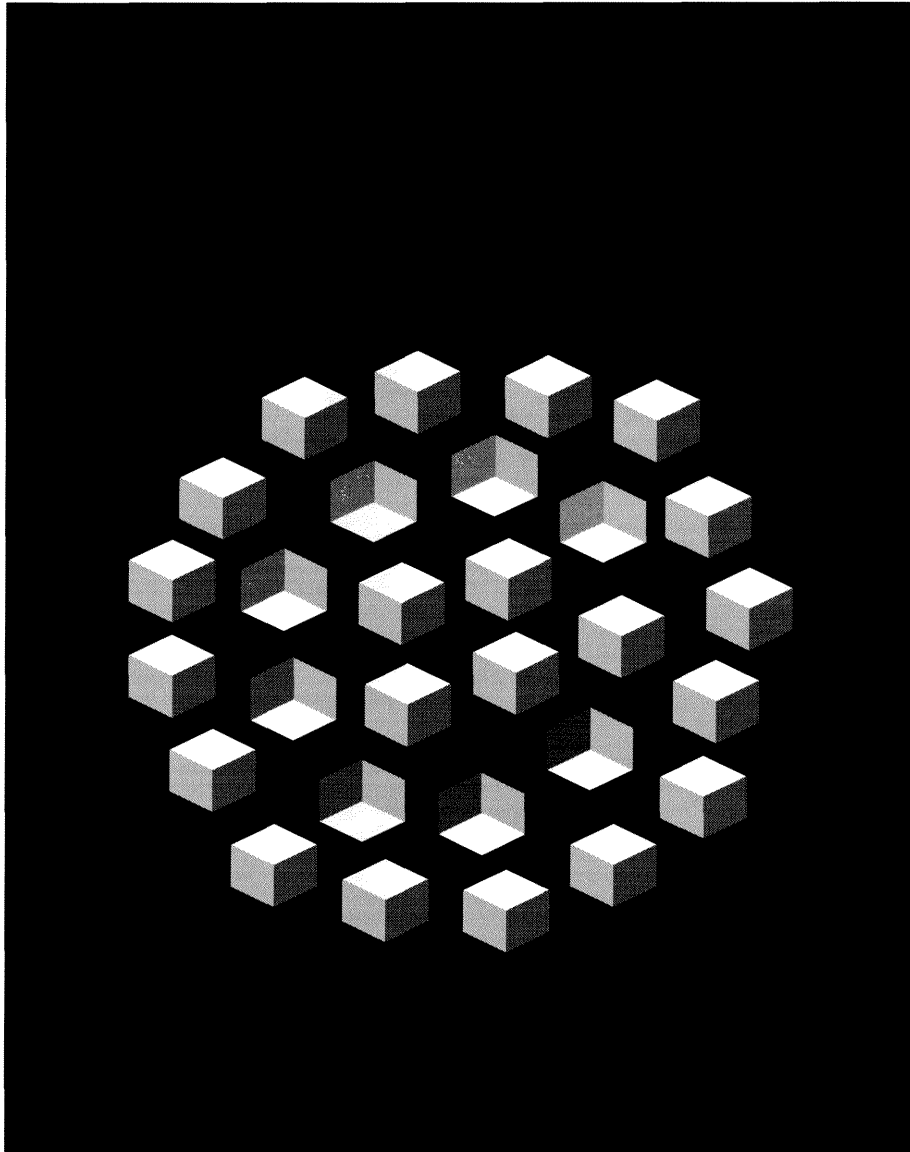


Figure 4.7: Upside-down cubes, forming a “C” pattern, are embedded among upright ones. All the upright cubes are identical in luminance pattern. The darkest side of the upside-down cubes, however, varies from light to dark, while the other two sides are held constant. Which of the upside-down cubes has a luminance pattern that is identical to the upright ones? After you have made your choice, turn the figure upside-down and make your choice again. Do your choices agree?

“C,” subjects consistently choose the fourth, or even the third, one. We suggest that a lighter cube than the correct one is chosen because the darkest side of the upright cube, interpreted as a 3-D shape, is perceived as lighter than its true luminance, due to the discounting of luminance as being a result of shading rather than reflectance. On the other hand, discounting occurs to a lesser degree for the upside-down cubes because it is less readily interpreted as a 3-D shape by the visual system (See Figure 4.8A). When Figure 4.7 is viewed upside-down, this illusion is destroyed. In this case, some subjects will choose the correct answer, while others will actually choose the next darker cube, the sixth cube, which is consistent with the explanation that the luminance differences of the cubes making up the “C,” now being upright and interpreted as 3-D shapes, are discounted to an extent.

This static-view demonstration also supports our hypothesis that pop-out is based upon reflectance computation (Figure 4.8B). Pop-out asymmetry is also seen (Figure 4.8C) when flat patterns that do not have any perceptual 3-D shape are used to create displays with reflectance maps equivalent to those depicted in Figure 4.8B.

Our results suggest that early vision processes compute some aspects of apparent reflectance together with 3-D shape. Fast and parallel discriminations are better explained on the basis of apparent reflectance than that of perceptual 3-D shape or brightness (See Figure 4.9). Perceptions of 3-D shape and reflectance are closely related (Metelli, 1974; Bergstrom, 1977; Gilchrist, 1979; Knill & Kersten, 1991; Adelson, 1993); our experimental results show that this relationship begins during the very early stages of visual processing.

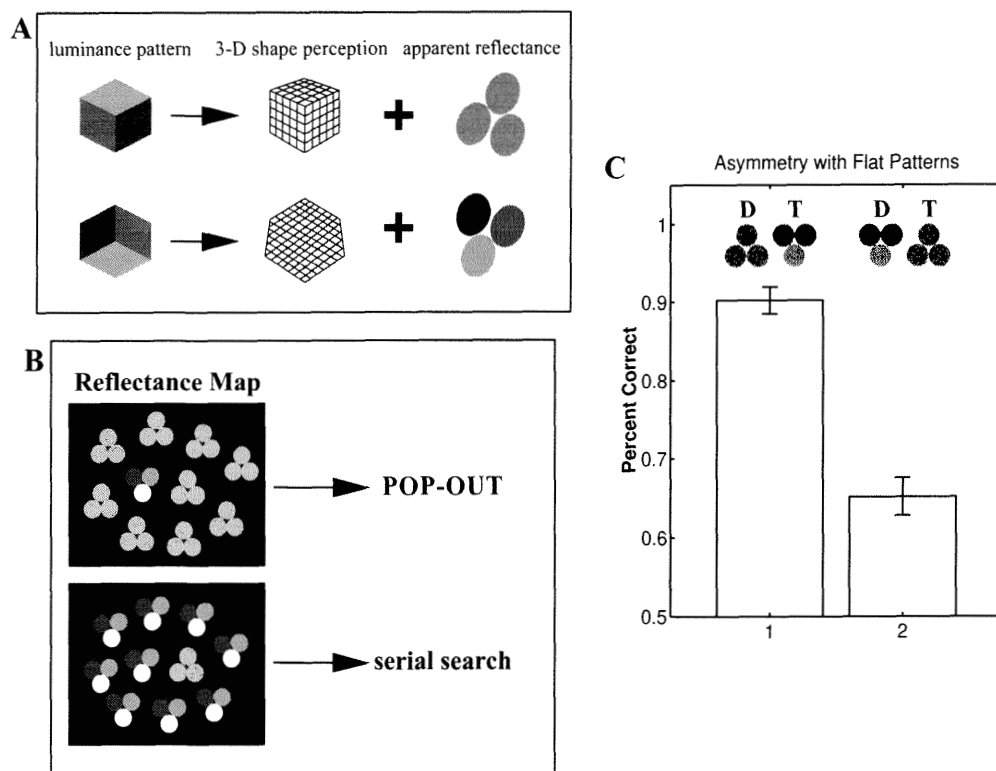


Figure 4.8: (A) We suggest that the top-lit, upright cube pattern promotes 3-D shape perception by early vision processes more so than the upside-down one. Such a difference in shape perception would imply a difference in reflectance perception. (B) A multiple-reflectance patch would pop-out from a background of single reflectance, while a single-reflectance patch would be difficult to spot in a multiple-reflectance background. (C) Stimulus displays that have reflectance maps equivalent to those on the left, but no perceptual 3-D shape, result in a similar asymmetry in performance under controlled experimental conditions. Data was taken from 2 naive subjects at a display duration of 120 msec.

Scheme for 3-D Pop-out

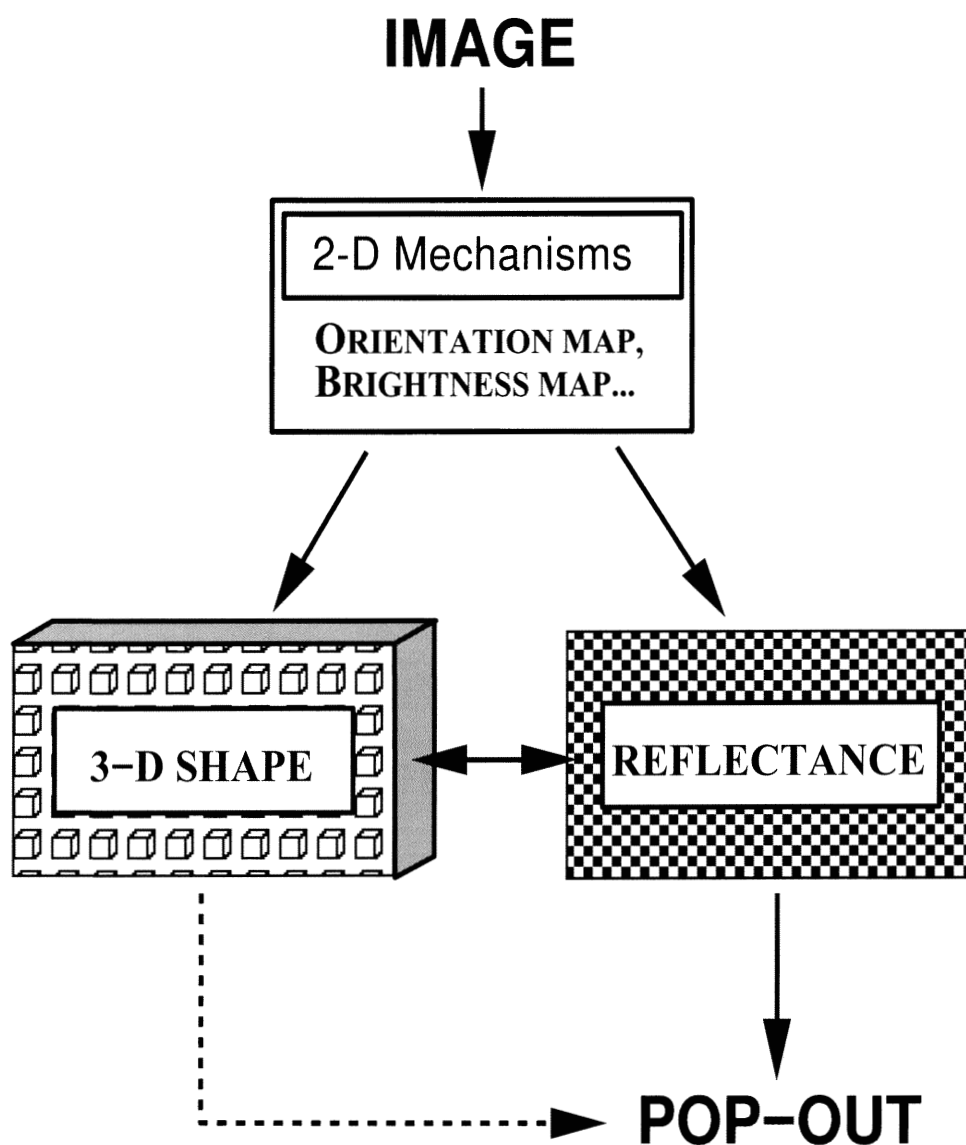


Figure 4.9: This diagram summarizes schematically our proposed scheme for 3-D pop-out. While we do not completely rule out pop-out arising directly from 3-D shape, our results suggest that this is weak compared to the pop-out that arises from reflectance differences.

Chapter 5 Direction of Lighting - Where is the Sun?

More precisely, the question should be, “Where, do we think, is the sun?” The two columns of bubbles in Figure 5.1 are actually identically shaded, with one column being the 180-degree rotation of the other. The left column is typically perceived as being convex, while the right column is typically perceived as being concave. The contrasting percepts may be explained by the fact that our visual system has a strong tendency to assume that light is coming from above.

Ramachandran (1988) demonstrated, using smoothly shaded “bubble” stimuli like the ones in Figure 5.1, that this assumption of lighting-from-above is used also by grouping mechanisms. Lighting from above is apparently also assumed by early vision processes for interpreting fast and in parallel shaded, 2-D renditions of 3-D scenes. This has been shown using 3-D pop-out tasks by Braun (1990, 1993) and Kleffner & Ramachandran (1991), with shaded bubble stimuli, as well as by Enns & Rensink (1991) and Sun & Perona (1996c) with shaded cubes stimuli. Moreover, our previous work shows that when the distractors, but not the target, can be interpreted as convex and lit-from-above, pop-out occurs. But what exactly does lit-from-above mean? Does the visual system prefer similarly all lighting directions that are above the horizon, or does it have a more specific preference? This is the questions we wish to address with the following experiments.

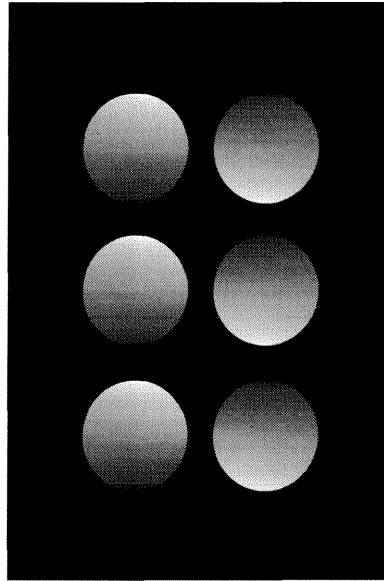


Figure 5.1: Shaded bubbles are typically interpreted as 3-D shape with the assumption that lighting is coming from above.

5.1 Methods

5.1.1 Method 1

Experiment 1 followed the same experiment procedure as that described for Experiments 1 & 2 of Chapter 4 (See Section 4.1).

5.1.2 Method 2

Images were generated on a Silicon Graphics Indigo². Stimulus screens were viewed at a distance of 20 inches, with the vertical height of the screen subtending 30 degrees of visual angle. Each stimulus patterns spanned approximately 1 degree. Items were displayed on randomized depth planes so that the target could not be discriminated based solely on disparity information.

We used a 2afc stimulus onset asynchrony (SOA) paradigm with masking. Stimulus screens were composed of 24 smoothly shaded circles that are typically interpreted

as 3-D convexities lit from a direction parallel to the shading gradient. Background distractors were shaded in one orientation, thus simulating a consistent lighting direction, while the target pattern, present during 50% of the trials, was shaded to simulate illumination from the opposite direction. The lighting direction, as defined by the shading gradient of the distractors, was held constant within a block and varied systematically between blocks. 12 lighting directions were used, spanning the 360-degree space evenly. Exactly vertical lighting from above was designated as 0 degree. Left-lighting was designated as positive degrees from the vertical, and right-lighting was negative degree values (See Figure 5.2).

Data was collected from 12 subjects, 6 right-handers and 6 left-handers, using a staircase method. The staircase method converges at 67% duration (MFV) within each block was used to estimate 66.67% accuracy performance, and is used as the criterion for comparison between lighting orientations. Each block consisted of 60 to 80 trials, with approximately 1000 trials presented during each session.

5.2 Experiments

5.2.1 Experiment 1 - Shaded Corners

In Experiment 2 of Chapter 4, all the distractor patterns were light on the left and dark on the right, and all the target patterns are light on the right and dark on the left. In Condition 1 (Fig. 5a), for example, the 3-D distractors could be interpreted as being lit from the left, and a significant improvement in performance was seen over the condition in which both target and distractors were flat. We collected data using right-lit distractors and left-lit targets as well (See Figure 5.3). If the visual system does not have a preference between lighting from the left and lighting from the right,

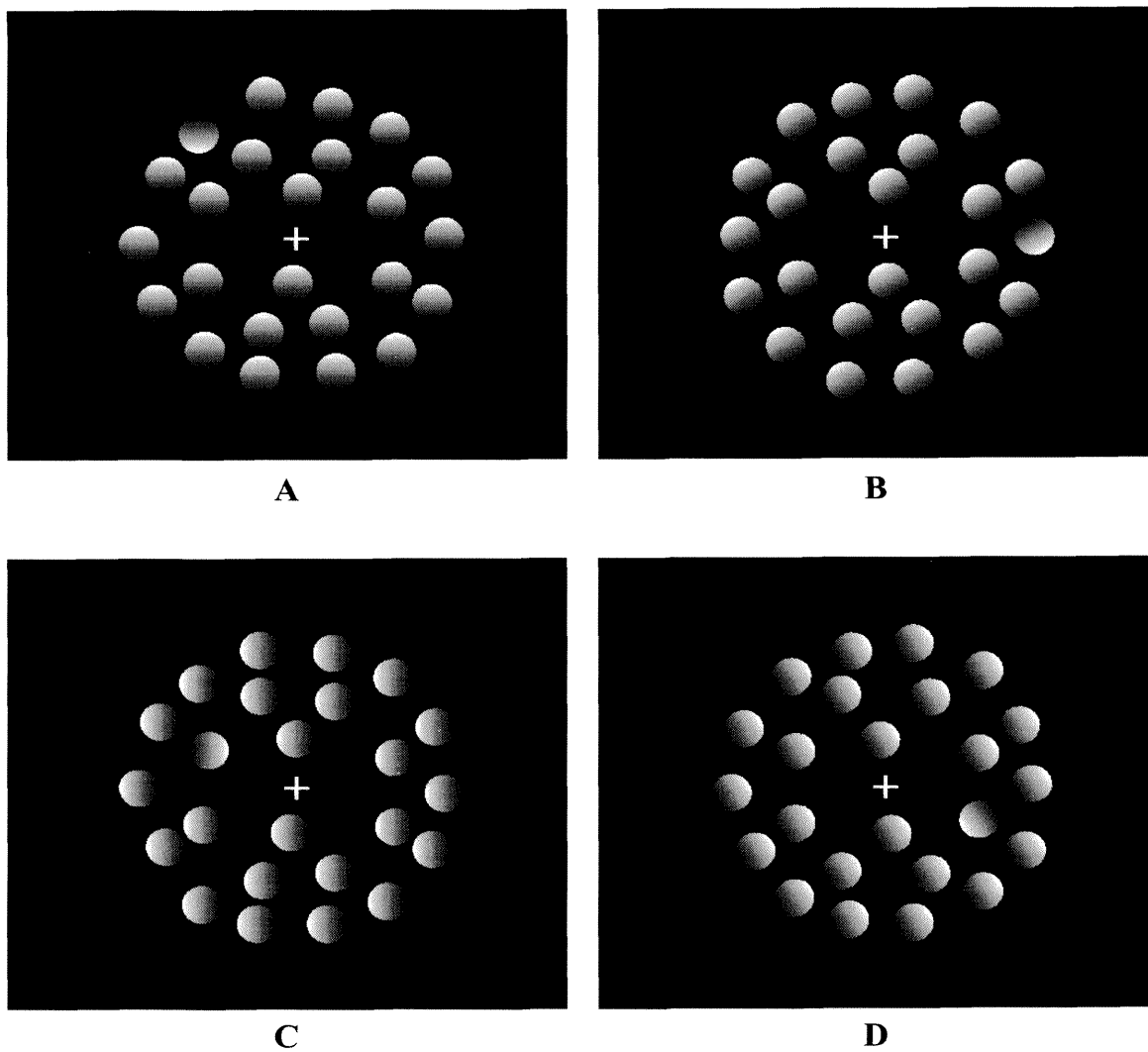


Figure 5.2: An example of a 0-degree stimulus screen is shown in A. Lighting from the left is denoted by positive angles from the vertical; a 30-degree screen is shown in B, and a 90-degree screen is shown in C. D shows a right-lit screen(-60 degrees).

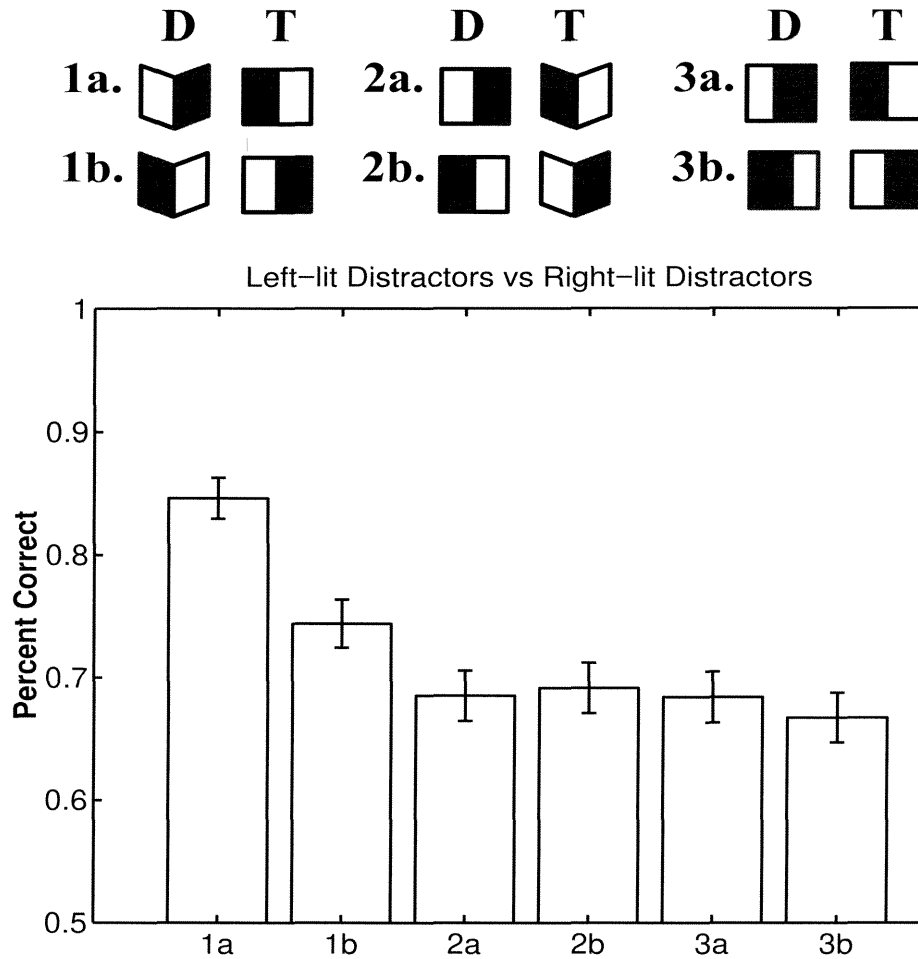


Figure 5.3: Results for 6 conditions displayed at a duration of 180 msec are shown. “A” conditions have distractors that are lit from the left, and “B” conditions have distractors that are lit from the right.

we would expect to see the same level of improved performance when 3-D shape is imposed using stereo disparity.

Figure 5.3 compares performance between conditions in which distractors were right-lit (B’s) and conditions in which distractors were left-lit (A’s). We see that:

1. Performances in Conditions 1B, 2B, & 3B do not differ significantly from each other.
2. While 2A does not differ significantly from 2B, and 3A does not differ significantly from 3B, we see a significant difference between 1A and 1B.

These results suggest that in early vision, lighting-from-left is preferred over lighting-from-right, in the sense that discounting of luminance differences occurs more effectively when the shading of the 3-D shape is consistent with the left-lit assumption.

5.2.2 Experiment 2 - Shaded Bubbles

Figure 5.4 shows data from one subject. The most frequently visited duration is plotted for each of 12 lighting orientations. 0 degrees represent vertical lighting from directly overhead, and 180 degrees represent lighting from directly below. The range between -90 and 90 degrees represent lighting from above the horizon. Negative degrees indicate lighting from the right, and positive degrees indicate lighting from the left. We see that, indeed, the visual system does not have a uniform preference for all lighting directions that are above the horizon. The most preferred direction of lighting is also not the directly overhead, 0-degree orientation. For this subject, the most preferred lighting direction appears to be between 30 and 60-degrees left of the vertical. Furthermore, a marked asymmetry is seen between the various left-lighting and right-lighting orientations: 30 degrees is preferred over -30 degrees, 60 degrees is preferred over -60 degrees, and 90 degrees is preferred to an even larger extent over -90 degrees.

We investigate more specifically this asymmetry between corresponding left-right angles of illumination by computing the mean difference between each of the three pairs of corresponding angles for all 12 subjects. The results are shown in Figure 5.5. The preference for left-lighting over right-lighting is apparent.

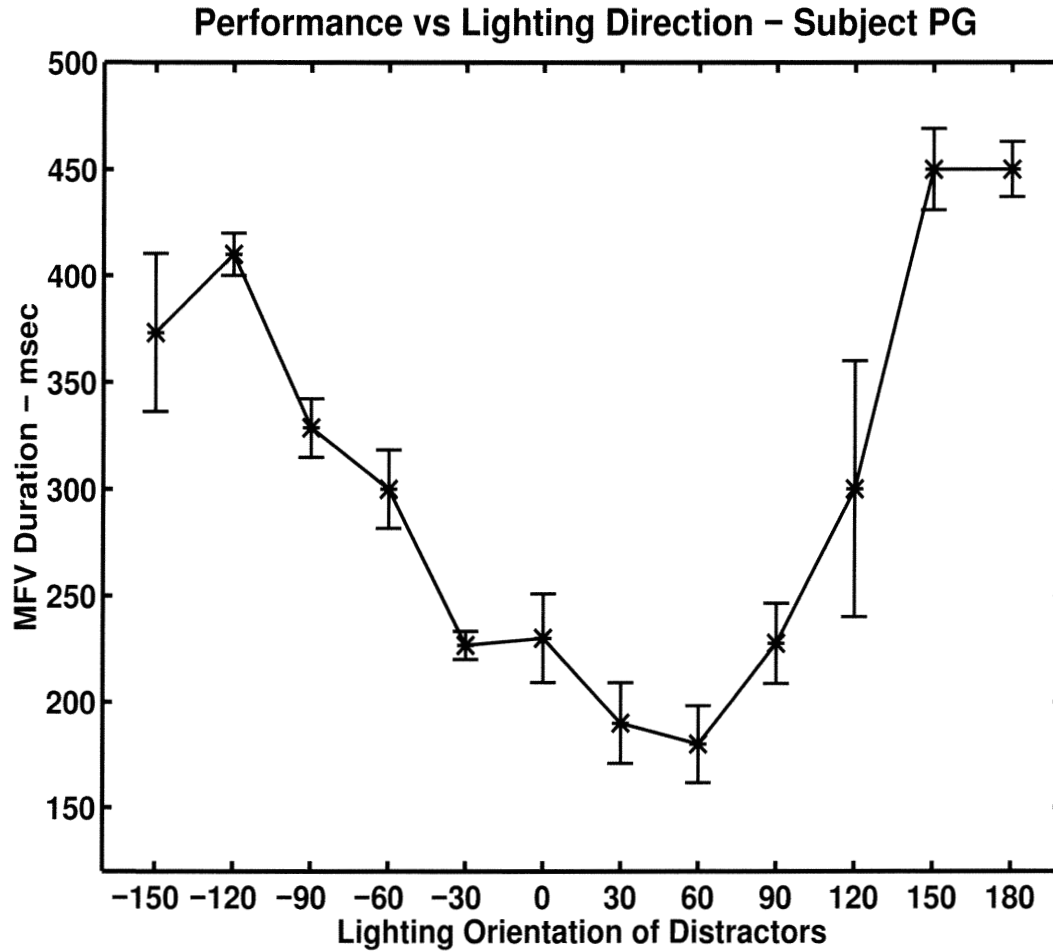


Figure 5.4: Lighting orientation, as defined by the distractors, is plotted on the x-axis; negative degrees indicate lighting from the right and positive degrees indicate lighting from the left. The SOA duration estimated for 67% accuracy is plotted on the y-axis.

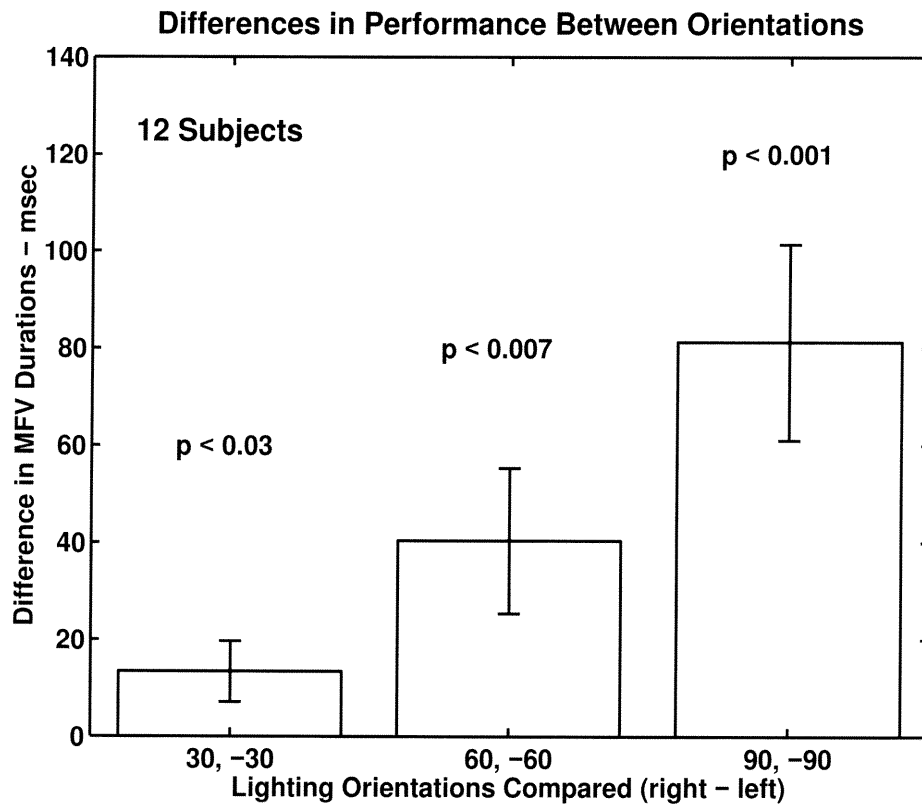


Figure 5.5: Each bar gives the average difference across 12 subjects between the performance for a right-lit condition with its corresponding left-lit condition. Results are given for the 30, 60, and 90-degree pairs. The p-value for significance is given above each bar.

5.3 Probable Causes

What is this asymmetry due to? We suggest that this left-right asymmetry may be related to one or more of the other more obvious left-right asymmetries that occur in human beings. We list four such asymmetries:

1. In-born handedness

Most of us are either right-handed or left-handed. Very few of us are exactly ambidextrous. Perhaps our handedness determines also our preference for left or right lighting. It is possible that, for whatever reason, right-handers prefer left-lighting and left-handers prefer right-lighting, or vice versa.

2. Handedness-related experience

Another possibility is that lighting preference, instead of being directly linked to the inborn handedness of a person, is the result of experience that is in turn related to handedness. For instance, when one arranges one's workspace, it makes sense to position a movable light source, or position oneself in relation to a fixed light source, such that one's dominant hand for manipulation does not cast a shadow upon the relevant areas of the workspace. A right-hander would be expected then to place his/her desk lamp on the left (Figure 5.6), resulting over the years, perhaps, in a preference for left-lighting. For left-handers, however, the situation is more complicated. While we would expect left-handers to do the opposite, thus resulting in a preference for right-lighting, we must consider the fact that left-handers live in a right-handers' world. Classrooms, for example, often have windows on the left instead of the right (Figure 5.7). We predict that, if handedness-related experience is what determines lighting preference, right-handers would prefer left-lighting, while left-handers, due to their mixed

experience, may exhibit a wider range of preferences, ranging possibly from a right-lighting preference to no preference to a less pronounced left-lighting preference.

3. Brain lateralization

Another biological asymmetry we have is the lateralization of our brain. Most people are left hemisphere dominant for language and analytical tasks, while the right hemisphere is believed to be dominant for spatial-perceptual tasks for most people. Since lighting convex scenes from the left would result in higher contrast, and therefore increased information, in the left visual hemifield, and the left visual hemifield is processed by the right hemisphere, we suggest that perhaps those of us who are right hemisphere dominant for spatial-perceptual tasks would also prefer left-lighting to right-lighting (See Figure 5.8).

4. Reading direction

While most languages are written and read from left to right, there are some languages, like Chinese and Hebrew, that are written and read from right to left. We propose this asymmetry to be another possible explanation for the left-right lighting asymmetry.

The above four hypotheses and predictions are summarized in Figure 5.9. We note that the hypotheses of inborn handedness, brain lateralization, and reading direction would all result in lighting preferences that are mirror symmetric for the left and right cases. The hypothesis for handedness-related experience, in contrast, predicts a left-lighting preference for right-handers, and a relatively weaker or mixed response for left-handers.



Figure 5.6: The working environment of a right-hander is usually set up with the light on the left.

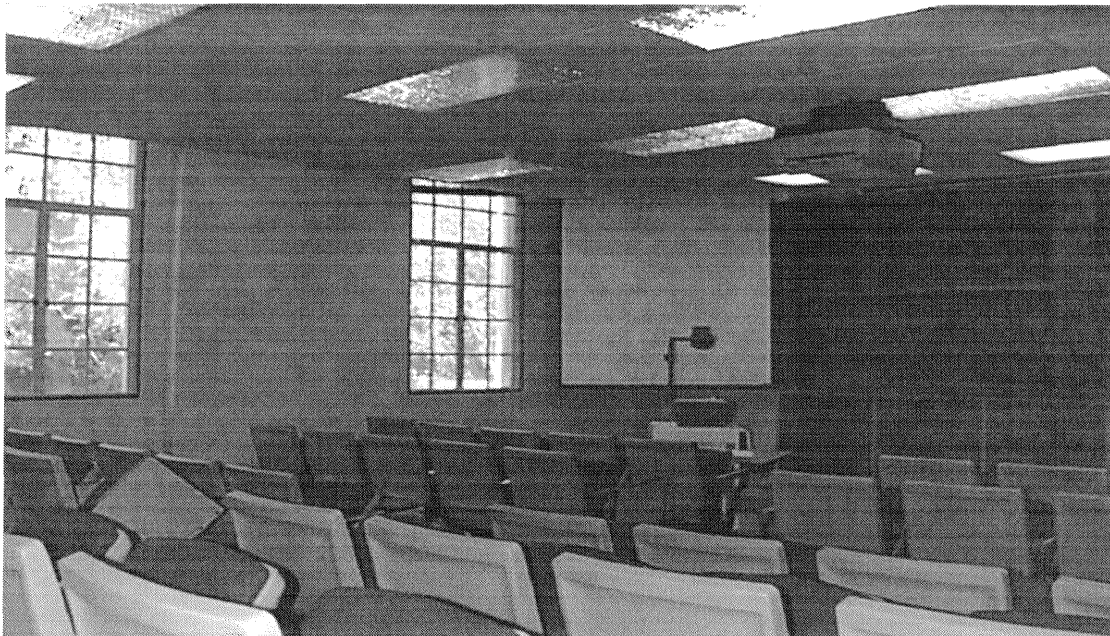


Figure 5.7: Because most students are right-handed, classrooms tend to be set up with windows on the left.

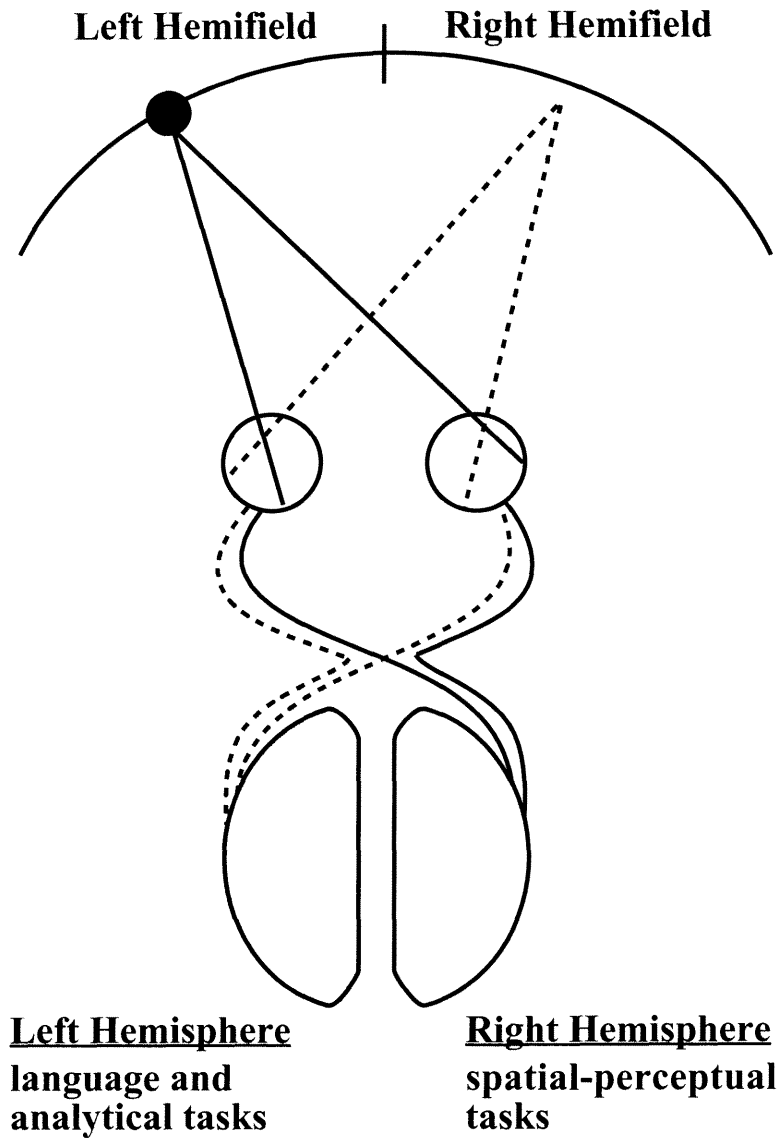


Figure 5.8: Another biological asymmetry we have is our asymmetry in brain lateralization. Most people are right-brain lateralized for spatial-perceptual tasks, while some are flipped with respect to the norm.









HYPOTHESIS	PREDICTION	
	Left	Right
Inborn Handedness		
Brain Lateralization		
Reading Direction		
Handedness Related Experience		

Figure 5.9: The predictions that follow from each hypothesis are summarized in this chart.

5.3.1 Lighting Preference and Handedness

We fitted a parabolic curve to the relevant portion of the psychometric curve of each subject. The lighting direction at which the minimum of the parabola occurs is used as an estimate of the subject's preferred lighting direction (Figure 5.10).

After obtaining an estimate of the preferred direction of lighting from each subject, we averaged the data according to the self-declared handedness of the subjects. The average preferred direction of lighting is shown for each group is shown in the bar graph of Figure 5.11.

Since handedness is not a binary characteristic, we also evaluated the relative handedness of our subjects using a standardized 10-item questionnaire developed by Oldfield (1971). Left-handedness is indicated with a negative score, while right-handedness is indicated with a positive score. The absolute value of the score is used as an indicator of the strength of handedness.

Figure 5.12 shows preferred lighting direction plotted against handedness. The correlation coefficient is 0.67. From Figures 5.11 & 5.12, we make the following

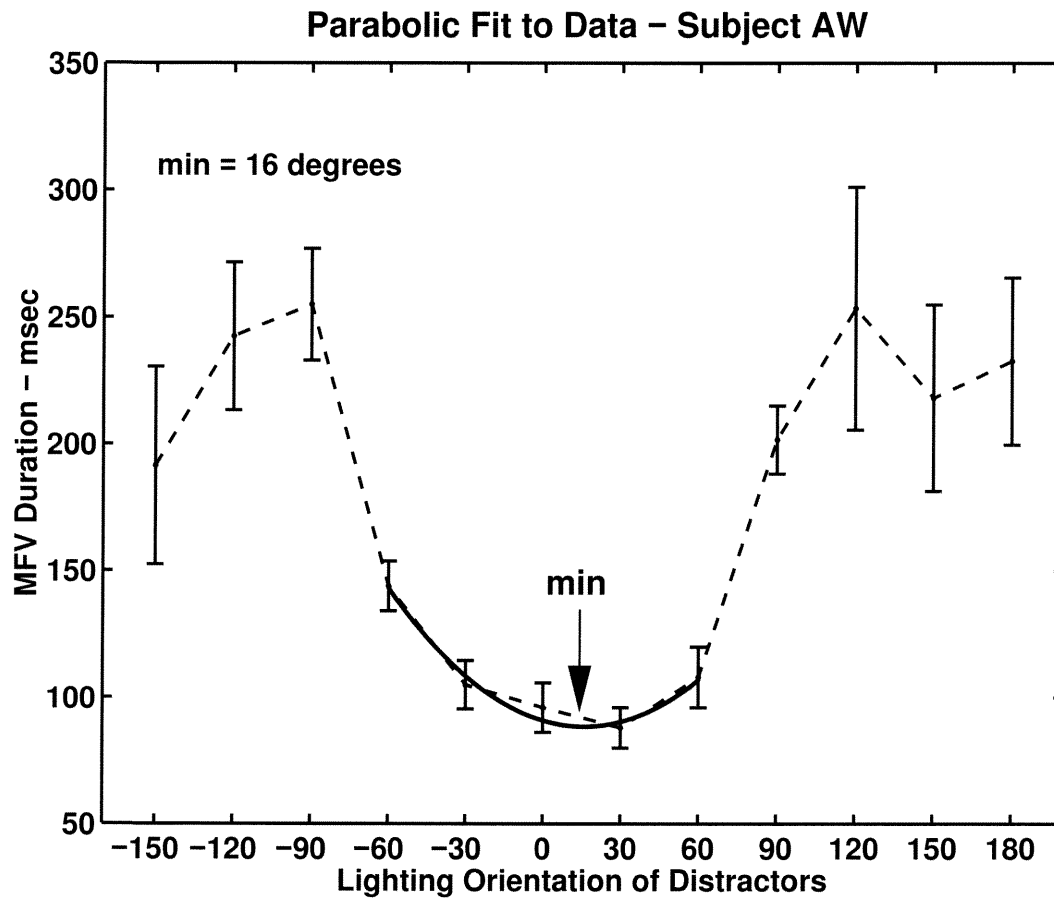


Figure 5.10: A parabolic curve is fitted to the relevant portion of the data of each subject. This graph gives as an example the data and the fitted curve for Subject AW. For AW, the preferred angle of lighting is 16 degrees left of the vertical.

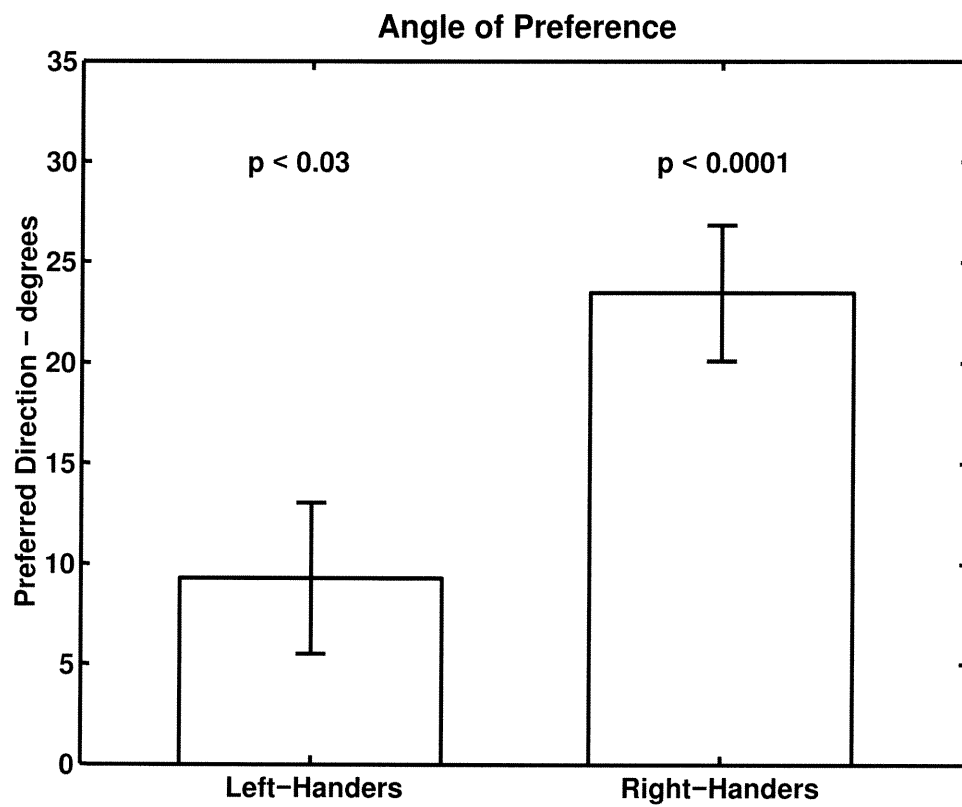


Figure 5.11: Preferred angle of lighting was averaged across right-handed subjects and left-handed subjects separately.

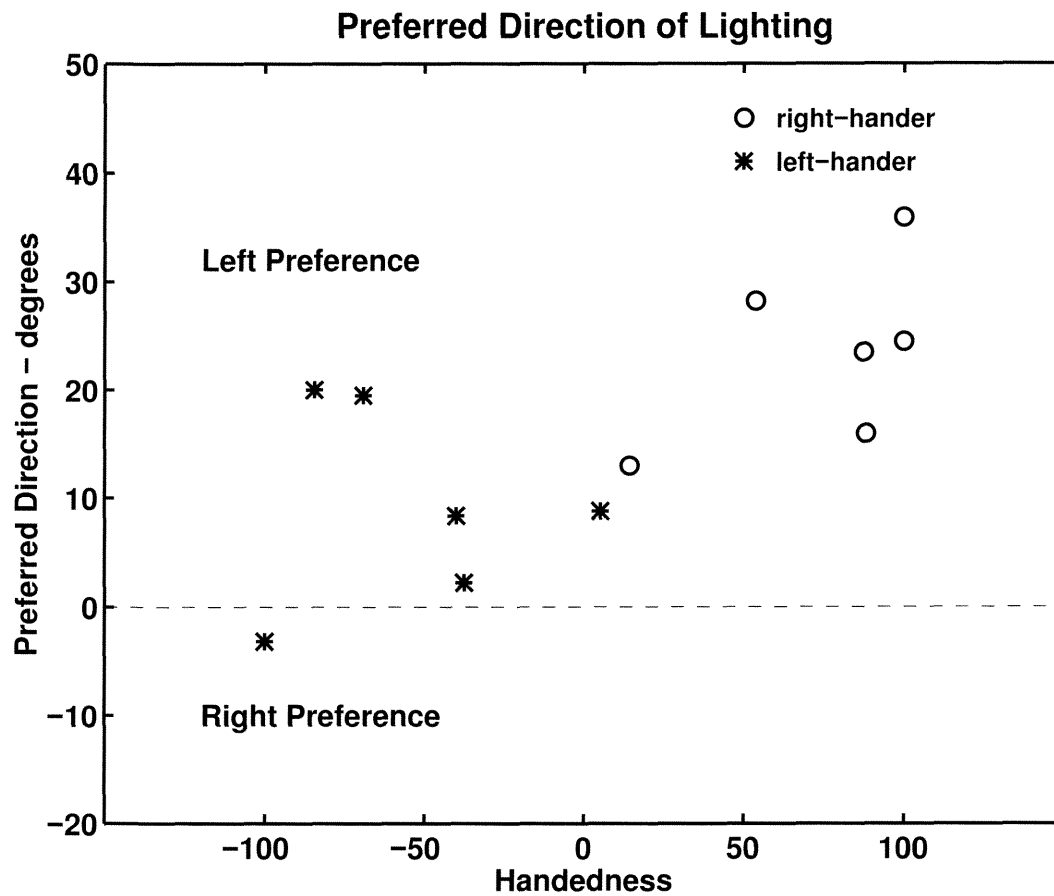


Figure 5.12: Preferred angle of lighting is plotted against a handedness score for each subject. The handedness score is derived from the Oldfield (1977) Inventory, a standardized questionnaire for evaluating handedness.

observations:

1. Preference of lighting direction is not strictly determined by inborn handedness. While all right-handers prefer left-lighting, so do most, but not all, left-handers.
2. Right and left-handers separate into two statistically distinct groups.
3. Right-handers as a group prefer a larger angle of left-lighting than left-handers.
4. There is a general trend of correlation between handedness and lighting preference.

5.3.2 Brain Lateralization and Lighting Preference

Next we wish to investigate the possible correlation between brain lateralization and lighting preference. We used the data collected for Experiment 2 to assay hemispheric dominance. We simply compared each individual subject's performance on trials in which the target appeared in the left hemifield to his/her performance on trials in which the target appeared in the right hemifield. We included only data for which the display duration was 180 msec or less.

Figure 5.13 shows the preferred visual hemifield plotted against handedness. As expected, the vast majority of the subjects are right-brain lateralized for our task, resulting in better performance for the left visual hemifield than the right. Our results of find only two subjects, both left-handers, who are left-brain lateralized for this task are consistent with the statistics given in the literature: approximately 30% of left-handers and approximately 5% of right-handers display brain lateralization that is flipped with respect to the norm (Rasmussen & Milner, 1977).

Our next figure (Figure 5.14) plots brain lateralization against lighting preference. We find that, while all subjects who are right-brain lateralized for this task prefers

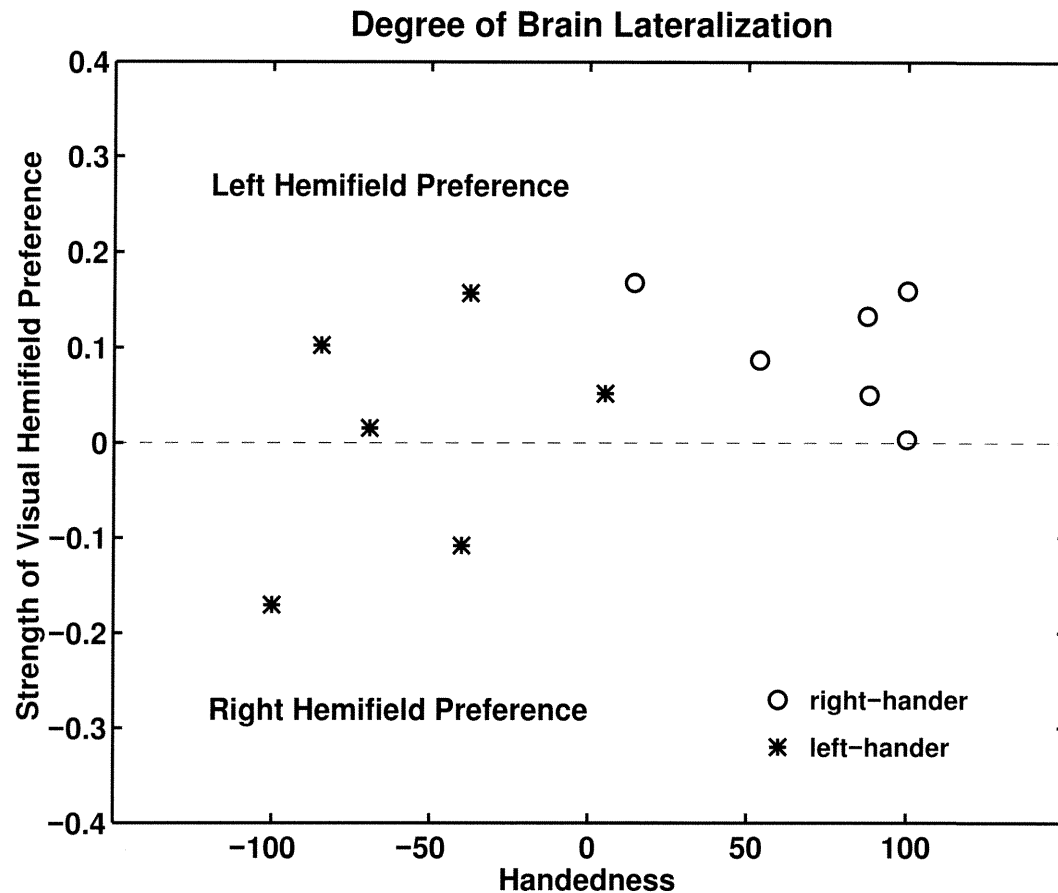


Figure 5.13: Handedness is plotted on the x-axis, and difference in performance between target present in left-hemifield trials and target present in right-hemifield trials is plotted on the y-axis.

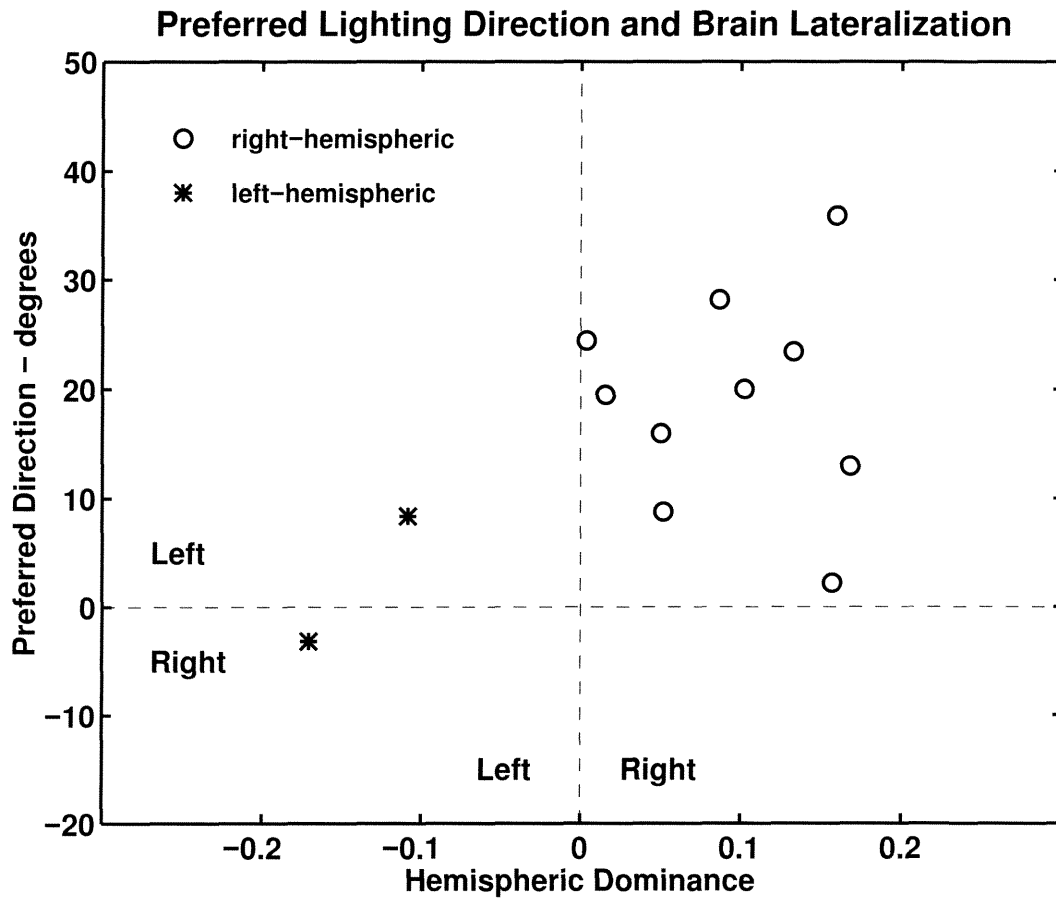


Figure 5.14: Preferred direction of lighting is plotted against hemispheric dominance for our 3-D pop-out task, as determined by hemifield preference

left-lighting, among the two subjects who display left-brain specialization, one prefers left-lighting and the other prefers right-lighting. While the data for the right-brain lateralized subjects suggest a correlation between brain lateralization and lighting preference, due to the small number of left-brain specialized subjects who participated in our experiment, it is difficult to make a conclusive statement.

5.4 Discussion

Early vision 3-D mechanisms make *a priori* assumptions about direction of lighting that are not discounted even when over-all shading patterns and stereo cues indicate the contrary. Lighting-from-above is preferred over lighting-from-below, and our results thus far suggest that lighting-from-left is preferred over lighting-from-right.

More specifically, our results show that the most preferred direction of lighting for the vast majority of the population is top-left, at around 25 degrees from the vertical. Right-handers show a stronger preference and a larger angle of preference than do left-handers. Among the hypotheses that may account for this left-lighting preference, we believe that our data most strongly supports the hypothesis of handedness-related experience.

While we have only just now begun to quantify this preference for top-left lighting, we suggest that artists, either consciously or subconsciously, have been influenced by this fact for hundreds of years. In a survey of 225 paintings taken at random from the Louvre, the Prado, and the Norton Simon Museum, two naive subjects classified $77\% \pm 0.55\%$ of the paintings as being lit from the left ($p < 0.05$). These subjects also estimated, using a protractor, the angle of lighting direction. Only the component of the angle that is within the plane of the painting was considered. The angles

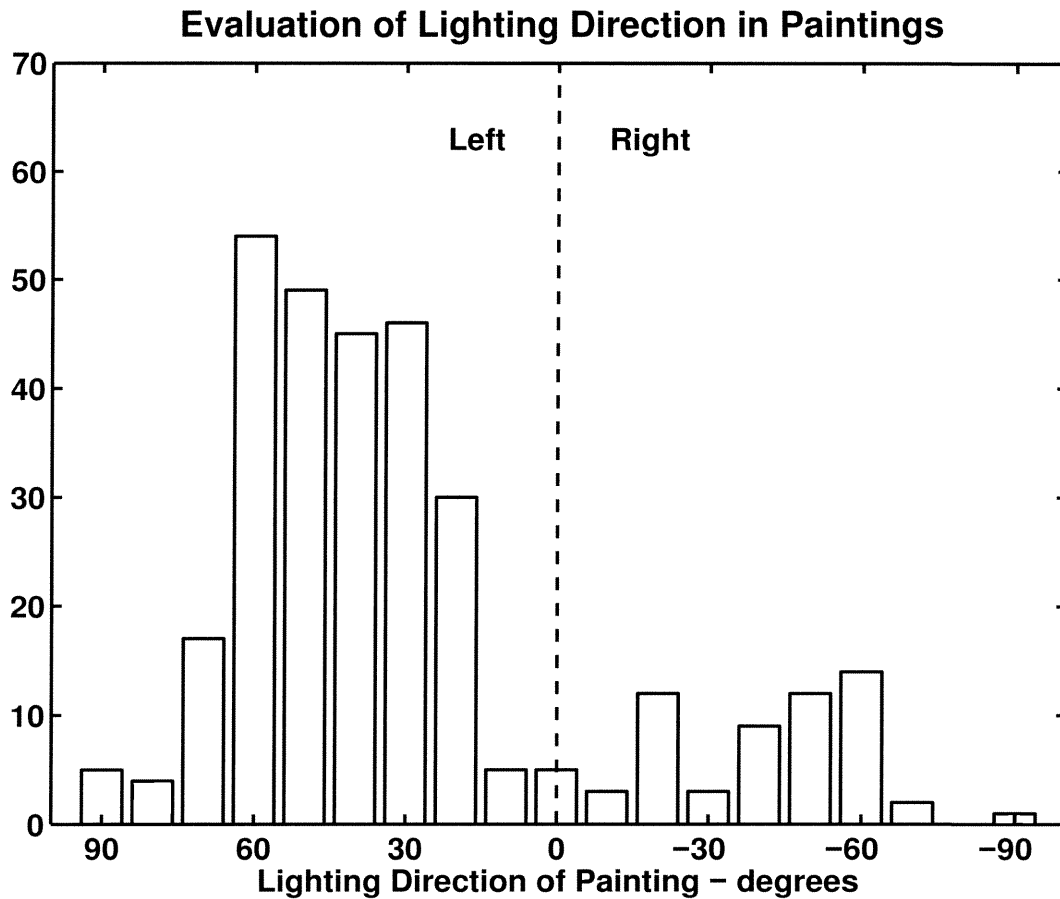


Figure 5.15: The lighting directions of paintings from 3 major museums are determined by two naive subjects. Results show a marked tendency for artists to choose top-left lighting.

of lighting as determined by the two subjects are histogrammed in Figure 5.15. The result shows that artists choose most often a lighting direction that is typically between 30 and 60 degrees to the left of the vertical. In fact, a survey of computer graphics images from the net reveals that 3-D icons and logos also show a marked tendency to be rendered as to suggest top-left lighting (Figure 5.16).

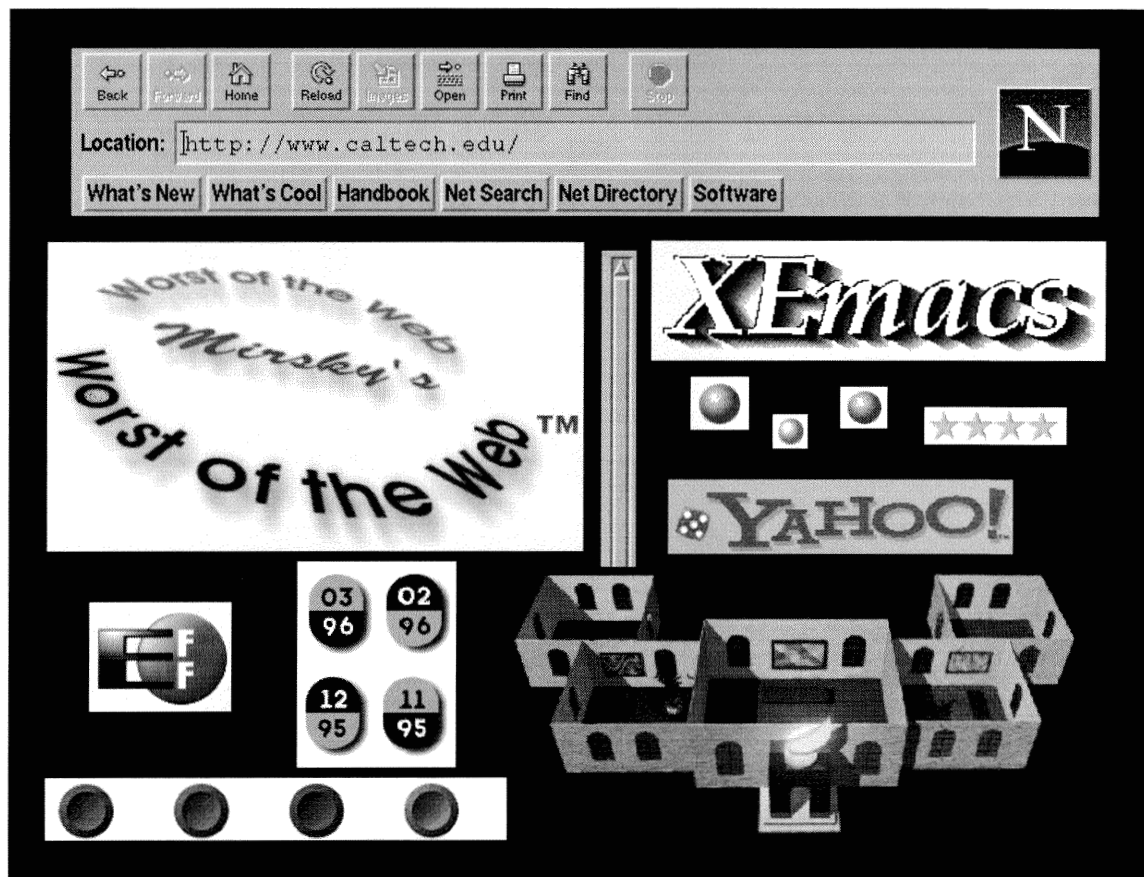


Figure 5.16: 3-D icons and logos found on the net are almost exclusively rendered with top-left lighting.

Chapter 6 Stereo and Shading

In Chapter 5 we addressed the issue that 2-D shaded images are intrinsically ambiguous, and the visual system, in order to arrive at an unambiguous 3-D interpretation, employs the assumption of top, or more precisely, top-left, lighting. Early vision processes seem also to have a preference for convex shapes over concave ones. This additional preference is suggested by asymmetric pop-out behaviors observed by Braun (1990, 1993) and Sun & Perona (1993, 1996c). (See Chapter 3). Both these assumptions are illustrated by the target-distractor reversal task shown in Figure 6.1. When the distractor patterns are perceived as convex cubes which are showing their top sides, as in Figure 6.1A, Conditions 1 & 2, a dark-topped target is easy to spot among light-topped distractors. However, a light-topped target is difficult to spot among dark-topped distractors (Figure 6.1B). If Figure 6.1 is viewed upside-down (Conditions 3 & 4), both sets of distractor patterns can be perceived as convex cubes that are showing their bottoms. In this case, results show that a light-bottomed target is significantly easier to find than a dark-bottomed target (Figure 6.1C), presumably because early vision mechanisms preferentially deal with distractors that can be interpreted as convex and top-lit. Note, however, that these "inverted" cubes take longer to process, requiring a display duration of 200 msec, as opposed to the 130 msec display duration for the "normal" cubes. Similar results from reaction time experiments were reported also by Enns & Rensink (1990).

What happens, however, if stereo disparity information is available, leaving no doubt as to the 3-D shape interpretation? Do early vision processes then recognize

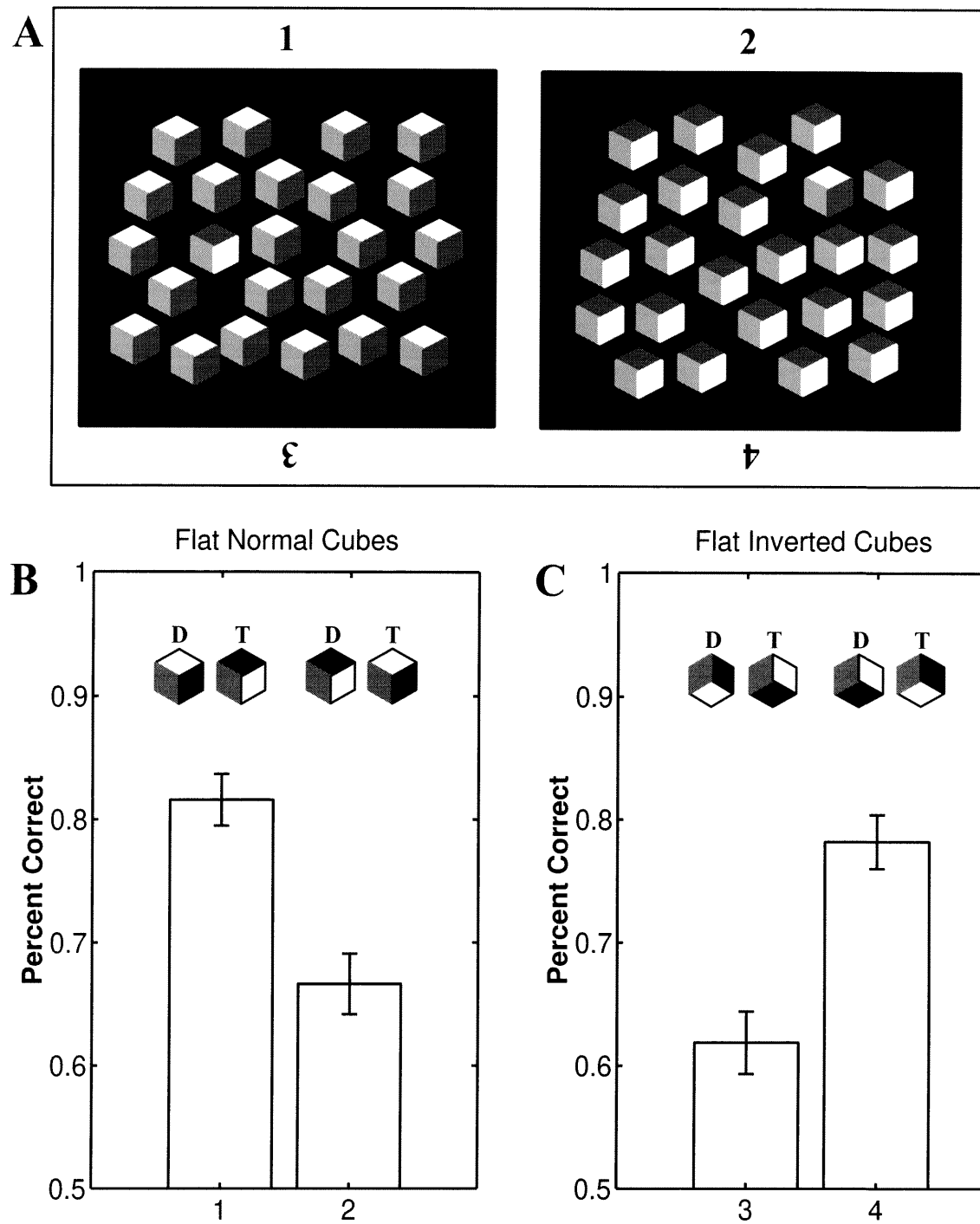


Figure 6.1: (A) Displays similar to those shown for Conditions 1 & 2 were shown at a duration of 130 msec. An asymmetry in performance is observed (B). Conditions 3 & 4, displayed at 200 msec, also resulted in an asymmetry (C), but in the opposite direction.

these assumptions of lighting and convexity as redundant and consequently discards them?

In the following experiments, we will address the “strength” of these two assumptions individually by imposing 3-D shape via stereo disparity cues onto shaded patterns.

6.1 Methods

The following experiments were conducted with the same experiment procedure as that described for Experiments 1 & 2 of Chapter 4 (See Section 4.1). The only difference is in the micropatterns that were used. Instead of 2-panel and triptych micropatterns, we used shaded cube micropatterns similar to those displayed in Figure 6.1.

6.2 Experiments

6.2.1 Experiment 1 - Convex Shapes

We investigate the top-lighting assumption in Experiment 1 by forcing convex shape interpretation onto shaded cubes with the use of stereo disparity cues. The upright cubes, both flat and convex conditions, were displayed at a duration of 130 msec, while the inverted cubes were displayed at a duration of 200 msec. If the top-lighting assumption is unnecessary when there is unambiguous 3-D shape, we would *not* expect to observe asymmetries like those shown in Figure 6.1.

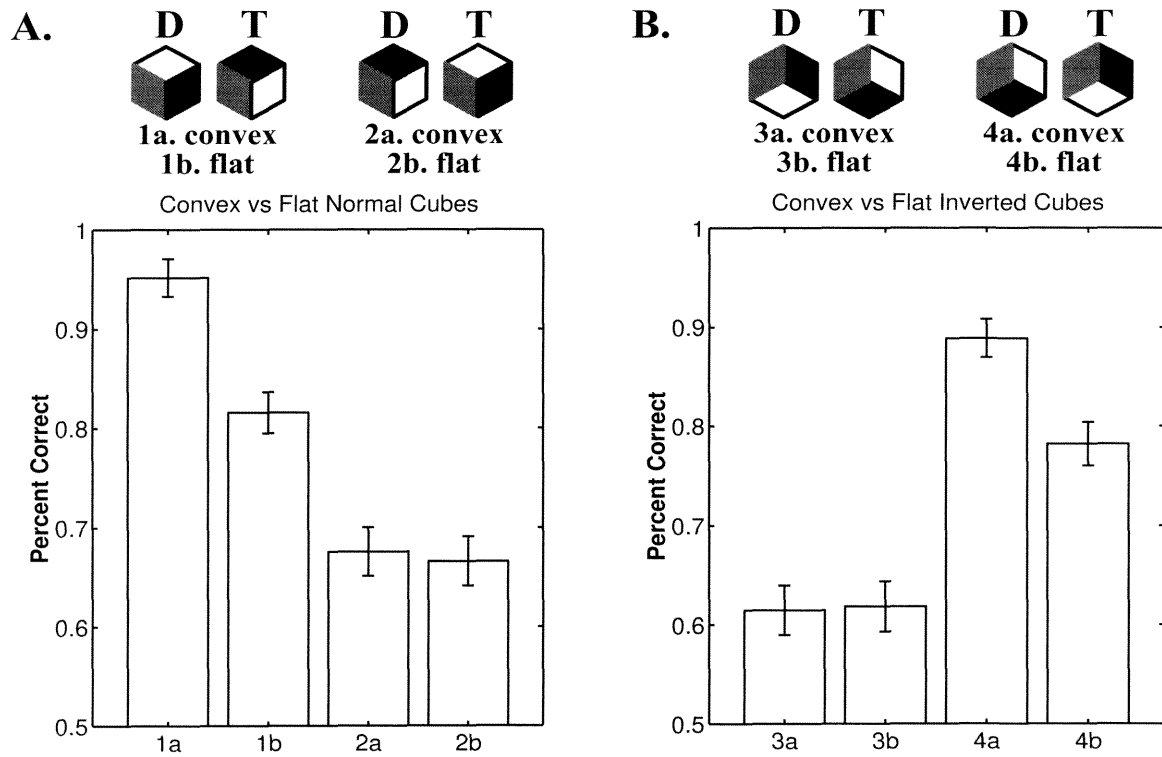


Figure 6.2: When 3-D shape is conferred using stereo disparity cues, the asymmetry in performance is enhanced for both the normal cubes (display duration = 130 msec) and the inverted cubes (200 msec).

Results

Results are shown in Figure 6.2. An asymmetry is seen between the convex conditions (Conditions 1a & 2a; Conditions 3a & 4a) as well as the flat conditions (Conditions 1b & 2b; Conditions 3b & 4b). We do not observe “rescue” from asymmetry.

It is possible no rescue was observed simply because shape from stereo disparity cues may require longer than the 130 msec display duration to establish, and the information is therefore discarded. We suggest that this is not the case, since a significant improvement can be observed when the easy flat conditions (Conditions 1b & 4b) were made convex by stereo (Conditions 1a & 4a). By imposing stereo disparity cues, not only was the asymmetry not rescued, it was in fact *enhanced*.

6.2.2 Experiment 2 - Concave Shapes

In this experiment we investigate the effects of imposing concave shape via stereo disparity cues. An identical experiment as Experiment 1 is conducted here, with the substitution of convex cubes by concave corners. If the concavity and convexity are equally preferred when there is unambiguous 3-D shape, we would expect to observe large effects, comparable to those observed in Experiment 1.

Results

Results are shown in Figure 6.3. For the upright conditions (Conditions 1 & 2), no improvement at all is seen when concavity is imposed using stereo disparity. A very slight impairment is seen when concavity is imposed in Condition 1. For the inverted conditions, we see a significant improvement when concavity was conferred in Condition 3, resulting in a partial “rescue” from asymmetry.

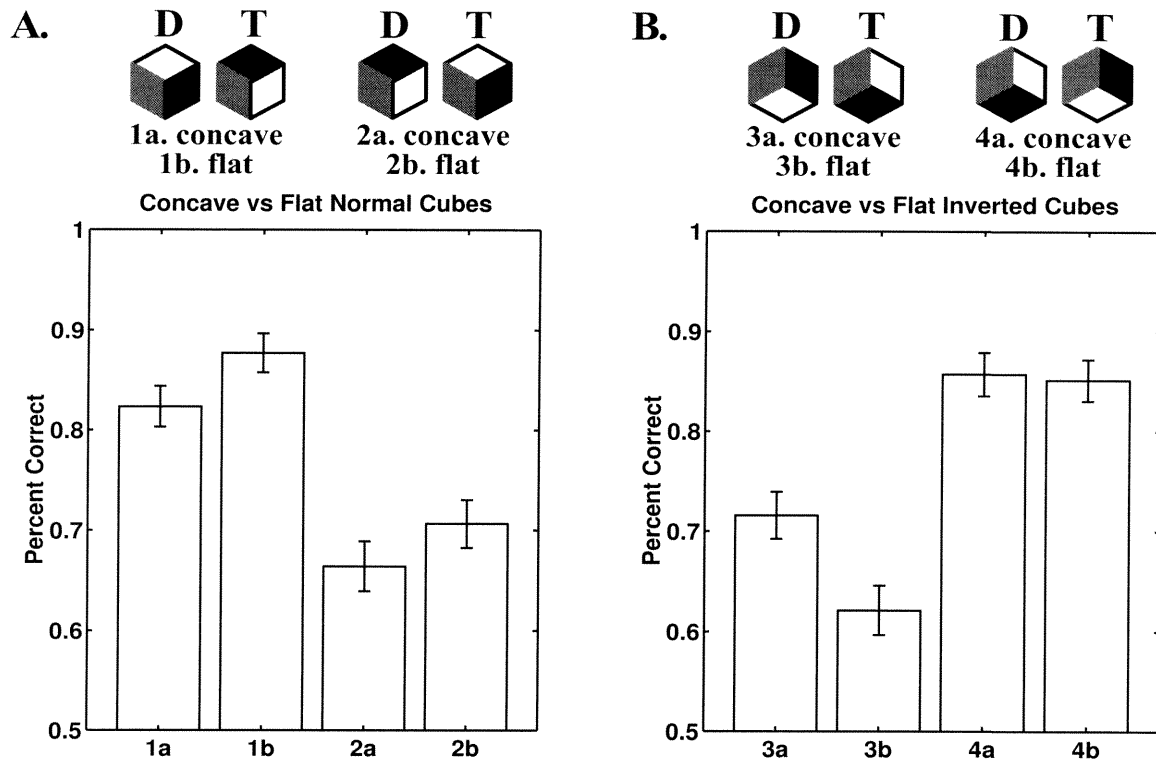


Figure 6.3: When concave 3-D shape is conferred using stereo disparity cues, the asymmetry is not significantly enhanced for the normal cubes (130 msec), and is significantly reduced for the inverted cubes (200 msec).

Overall, the large effects seen with the convex cubes over flat ones (Figure 6.2, Conditions 1 & 4) is not seen with concave corners. 3 out of 5 subjects reported that the concave percept was difficult to achieve. They reported that they could only see the cubes as concave through conscious effort.

6.3 Discussion

We suggest a possible explanation for the enhanced asymmetry seen in Experiment 1. In Chapter 4, we report experimental results consistent with the hypothesis that the primary cue for 3-D pop-out is reflectance, rather than 3-D shape or luminance. We proposed that a percept of 3-D shape leads to a discounting of apparent luminance, resulting in a more uniform apparent reflectance. Distractors that are interpreted as 3-D would result in a more uniform background that facilitates target detection, while distractors that do not promote 3-D interpretation would result in a “colorful” background of multiple reflectances, leading to difficult target detection (Sun & Perona, 1996a).

Both cases of asymmetry enhancement may be explained based on this hypothesis. In Conditions 1a & 4a, both the shading and the stereo cues of the distractors are consistent with a top-lit, 3-D interpretation. Luminance differences are discounted to an even larger extent than in the flat conditions due to strong shape interpretation from stereo cues. The targets, however, while also having a 3-D interpretation, have shading cues that are inconsistent with a top-lighting interpretation. While the distractors in Conditions 1a & 4a are interpreted as cubes of single reflectance, the targets are interpreted strongly as cubes of multiple reflectances, resulting overall in enhanced detection.

The lack of improvement seen in Conditions 2 & 3 when convexity was imposed using stereo cues further supports the notion that early vision 3-D mechanisms make a strong assumption for top-lighting. We suggest that the distractors of Conditions 2a & 2a are interpreted as cubes of multiple reflectances, while the targets are interpreted as cubes of uniform reflectance, resulting in conditions that are by no means any easier than the original flat conditions (2b & 3b).

These results are consistent with the proposal that early vision 3-D mechanisms make *a priori* assumptions about lighting direction that cannot be over-turned even when shape information is made unambiguous through the use of stereo disparity cues. Furthermore, we note that all significant effects of imposing convexity involved improvement of the easy flat conditions and not the hard ones. This suggests that an unambiguous convex percept is congruent with the default shape assumption; an assumption of convexity.

Experiment 2 shows that while imposing concavity can cause a slight impairment of an easy flat condition (Condition 1) and improvement of a difficult flat condition (Condition 3), unlike imposing convexity, it does not facilitate the processing of either of the easy flat conditions. This supports the idea that the early vision mechanisms involved here make an initial assumption for convex, not concave shapes. Overall, the effects are less prominent than those observed in Experiment 1, again suggestive of the notion that early vision 3-D mechanisms prefer convex shapes over concave ones.

The significant effects seen here are also in line with the hypothesis that easy discrimination is based upon reflectance differences. In Condition 1, concave disparity cues would lead to the interpretation of the distractors as concave corners of multiple reflectances. This would lead to more difficult detection. In Condition 3, the effect is larger, perhaps due to the longer display duration. We suggest here that imposing

concave disparity allows the distractors to be interpreted as being top-lit, thereby leading to facilitated performance.

The results of Experiment 2 and the reports given by the subjects suggest that early vision mechanisms process convex shapes more readily than concave shapes.

Overall, the results of Experiment 1 and 2 support the idea that early vision processes make default assumptions about lighting direction and shape that cannot be readily overturned even when 3-D shape is made unambiguous by stereo disparity cues.

Chapter 7 Influence of Contextual Information

The results from our pop-out experiments suggest the existence of mechanisms that can compute 3-D interpretations of shaded patterns composing a scene fast and in parallel. We also found that subject performance is highly correlated with reports of easy 3-D scene interpretation during short display durations. Is this 3-D pop-out based solely on local mechanisms, or can this process be influenced by global and/or contextual information? The following set of experiments investigate this question.

7.1 Methods

7.1.1 Experimental Set-Up

The same 2AFC SOA with masking paradigm was used for this second set of experiments. In Experiments 1 & 2, 6, 12, 18, or 24 items of display were used. In test experiments, items ranged in size from 0.9 to 2.6 degrees and were arranged according to size (See Figures 7.2, top row & 7.3, left). In addition, one condition involved a background that suggested the context of a room. The background was displayed throughout the duration of the experiment. In control experiments, all items had the same size (1.5 degrees of visual angle), and no background context cues were used. In Experiment 3, 3, 6, or 12 items were displayed at an eccentricity of 4.3 degrees of visual angle, with random jitter of up to 0.3 degrees. In test experiments, the items were displayed within a wall frame that has a 3-D interpretation. In control

experiments, the frame did not have a 3-D interpretation (See Figure 7.6). Both 2-D and 3-D frames were displayed statically throughout the experiment.

7.1.2 Data Analysis

Psychometric curves were fitted using the same method described in the Data Analysis section of Chapter 3. SOA durations necessary for obtaining 75% accuracy were estimated for both test and control conditions for each subject. Individual improvements were combined by weighted averaging to give the mean improvement and an associated standard error. Data is presented in terms of mean improvement of performance under test conditions with respect to control conditions. The probability that improvement is significantly greater than zero is given as a measure of confidence.

7.2 Experiments

7.2.1 Experiment 1 - Normal Orientation Cubes in Room Context

We see from Figure 7.1 that cubes of different sizes make a much harder task than cubes of all the same size, a size that is about the average of the cubes in the different-sizes condition. Necessary SOA duration for the largest display size increases more than 100 msec. In Experiment 1, we investigate the condition in which the cubes are of different sizes, but are arranged in an orderly size gradient.

In one display condition, the arrangement mimics the effect of cubes sitting on the ground, receding off into the distance (See Figure 7.2 top-left). Figure 7.2 (bottom-left) graphs the results in terms of improvement over control condition of same-sized

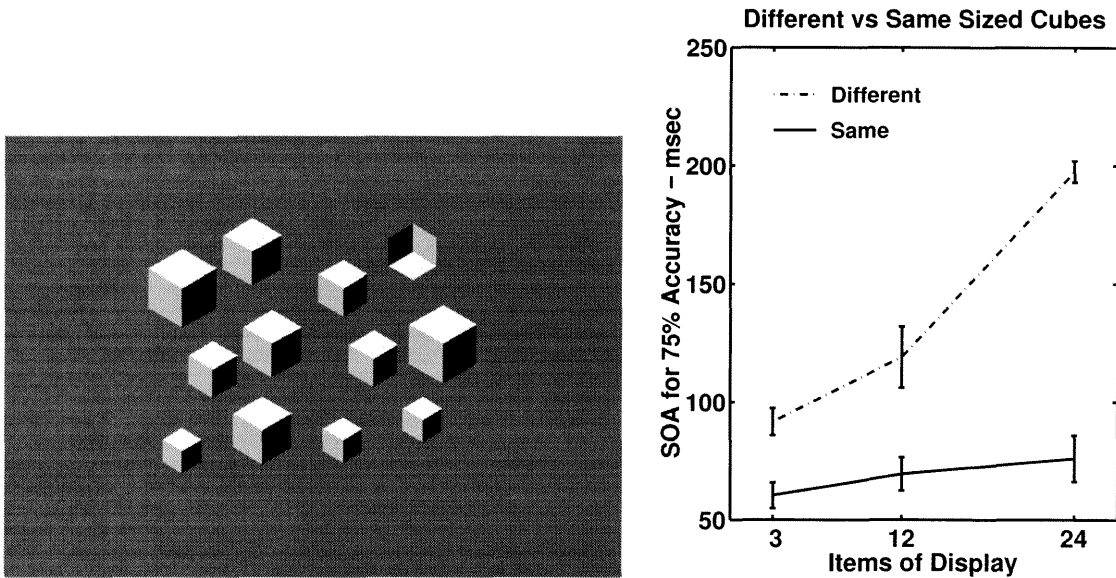


Figure 7.1: Performance of three subjects was tested using shaded cubes of different sizes (left). The results are shown in comparison to that of same-size shaded cubes on the right.

cubes, with degradation of performance represented as negative improvement. There is an impairment of 20 to 30 msec.

When the perspective cues were reinforced by a background room context (See Figure 7.2, top-right), however, we found an overall trend of improvement. The confidence levels of improvement for the four increasing display sizes are 93%, >99%, >99%, and 96% respectively. Subjects reported that the perspective sizing enhanced the 3-D percept and made the task easier. In contrast, for what we call the ceiling perspective with room context (Figure 7.3), which is a rather unusual viewing condition that does not fit with the apparent orientation of the shaded cubes, results compiled from 3 subjects show significant negative effects for the larger display sizes.

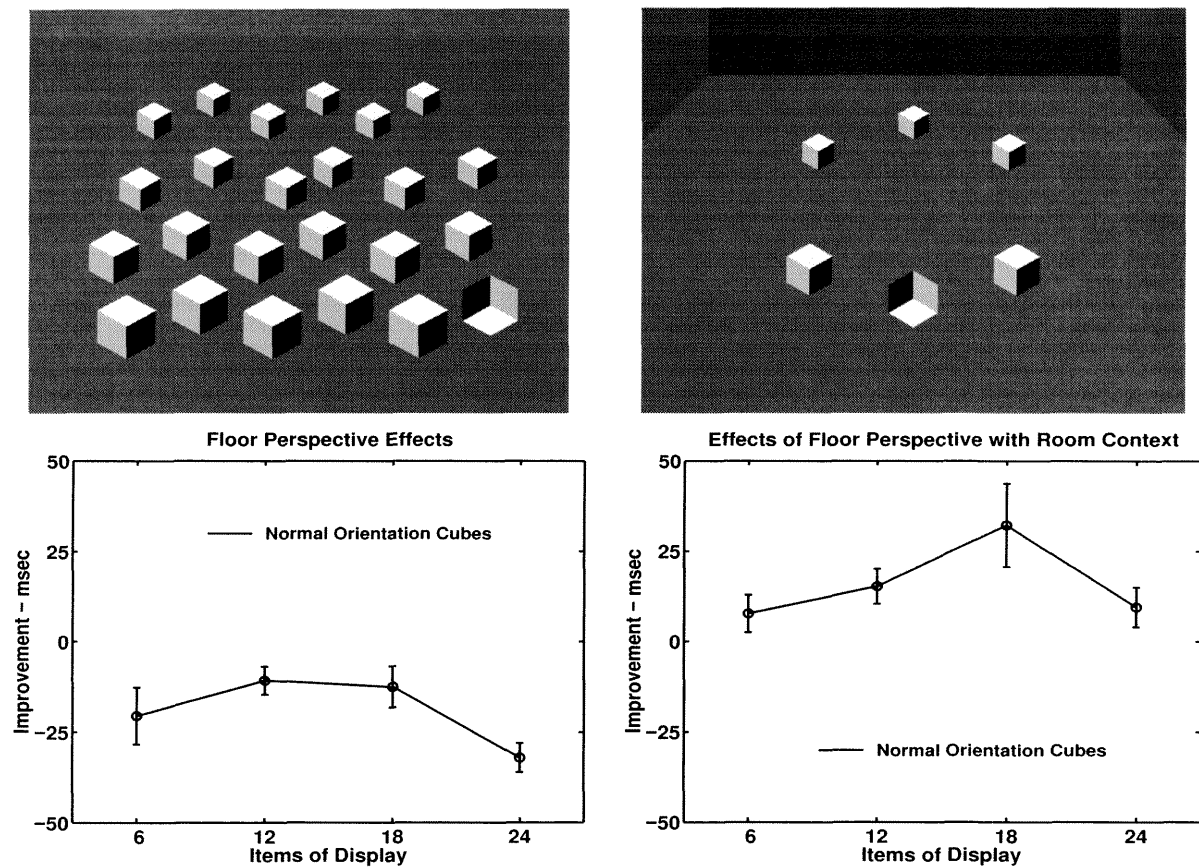


Figure 7.2: Top row shows normal orientation cubes arranged in floor perspective (left) and normal orientation cubes in a room context as well as in floor perspective (right). Bottom row shows the respective effects of these enhancements in terms of improvement. These plots reflect the data collected from 4 and 3 subjects, respectively.

7.2.2 Experiment 2 - Reverse Orientation Cubes in Room Context

Experiment 2 deals with the effect of contextual information on reverse orientation cubes. Three subjects were tested on the ceiling perspective only experiment, and two subjects were tested on the one that included a room context.

When only perspective is used, we see a generally insignificant effect, except for the 6-item case (Figure 7.4, left). When room context is added, however, the improvement becomes significant (confidence level at >99% for all display sizes), and

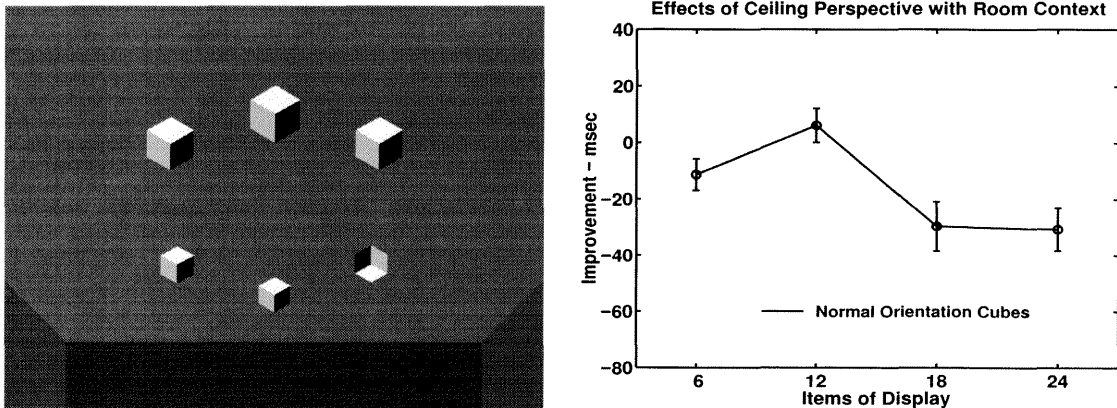


Figure 7.3: Normal orientation cubes were arranged in ceiling perspective and room context (left). The improvement of this condition over the standard task is shown on the right.

especially large for the two larger display sizes (Figure 7.4, right). Although we originally thought that these reverse orientation cubes would be best perceived as cubes hanging from the ceiling and therefore a floor perspective should hinder performance, some subjects reported that the floor perspective enhanced perception of the 3-D scene by allowing the stimuli that were previously difficult to interpret to be perceived as cubes balanced on a single vertex.

To test our hypothesis that the reverse orientation cubes might be best perceived as bottom-lit cubes hanging from the ceiling, we tested 3 subjects on stimuli consisting of reversed cubes arranged in a ceiling perspective, as shown on the left of Figure 7.5. Data was only collected for 3 display sizes: 12, 18, and 24 items. We see overall improvement that is correlated with subjects' reports of enhanced 3-D perception (Figure 7.5, right). The confidence levels of improvement for the three increasing display sizes are >99%, >99%, and 97%.

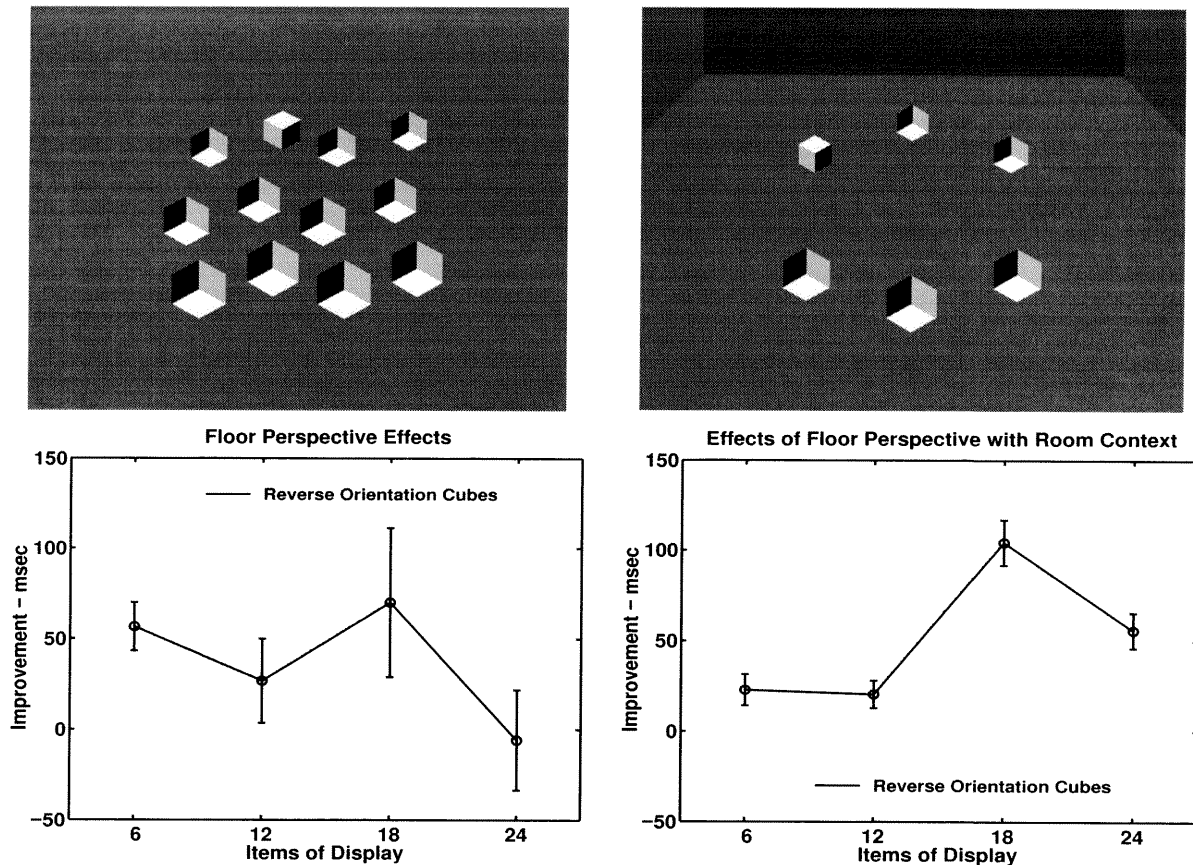


Figure 7.4: Reverse orientation cubes arranged in floor perspective (left) and with room context (right) is shown on the top row. Bottom graphs depict the respective effects.

7.2.3 Experiment 3 - Holes in Walls

In this experiment, we extend our investigation of contextual effects to a pattern other than the shaded cube. For a rotated Y-junction in a square, which may be interpreted as a hole, we ask the question: how does a context that has either a consistent or an inconsistent 3-D interpretation with respect to its embedded patterns affect performance? Figures 7.6A & B show the displays which have distractor holes that are respectively consistent and inconsistent with the "3-D" wall frame. For control, an analogous surrounding frame that has no 3-D interpretation was used (Figure 7.6C).

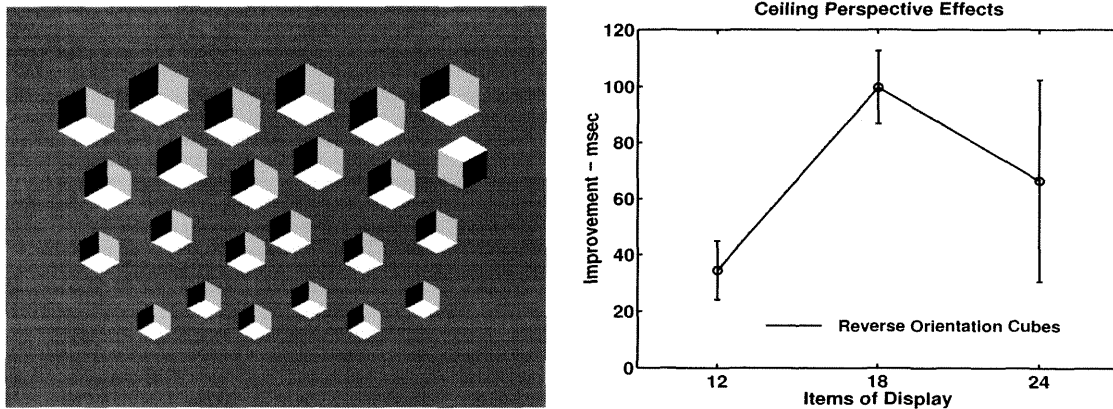


Figure 7.5: Reverse orientation cubes were also tested in ceiling perspective (left). The results are summarized in this graph in terms of improvement (right).

Data was collected from 4 subjects for the consistent 3-D frame experiment, and 3 subjects for the inconsistent 3-D frame experiment. Compared with the controls, the 3-D frame that was consistent with the distractor holes facilitated performance significantly, at a confidence level of $>99\%$ for every display size. There is also a trend for larger improvements to occur for larger display sizes (Figure 7.7, left).

The 3-D frame inconsistent with the distractor holes did not lead to statistically significant improvements or impairments (Figure 7.7, right). For the inconsistent frame case, some subjects saw the distractors as protruding cones, which would be consistent with the shading of the frame, instead of inconsistent holes. Other subjects saw the distractors as inconsistent holes only. We suspect that performance may have been facilitated for those who formed the consistent percept, but not for those who formed only the inconsistent percept. This dichotomy in perception might explain the large error bars.

7.3 Discussion

The results from all three sets of inducement experiments suggest that contextual information influences perception and performance. There are instances of improvement as well as impairment, with effects that are generally larger for the larger display sizes (Figures 7.1 - 7.3, 7.4 (left), 7.7 (right)). We believe that these results can be best understood as a combination of bottom-up textural effects and top-down expectation effects.

7.3.1 Textural Effects

Figure 7.1 shows a correlation between the disruption of textural uniformity and the break-down of perceptual pop-out. This result suggests the involvement of textural mechanisms alongside the fast and parallel processing of 3-D shapes.

Dramatic impairment of performance is "rescued" to a large extent, however, if the different-sized cubes, instead of positioned at random, are arranged according to size (Figure 7.2, left). The largest increase in necessary SOA is about 30 msec instead of over 100 msec. We suggest that this partial rescue is due to the fact that the cubes, when arranged according to size, give rise to a texture that is, at least locally, homogeneous. Background homogeneity has been shown to a large effect on search efficiency (Duncan & Humphreys, 1989; Wolfe *et al.*, 1992). The remaining impairment could be explained if the texture mechanism preferentially subserves uniform textures, or if some of these cubes are of a size that hinders discrimination, either too big or too small.

Not only can the texture gradient perform partial rescue, we see from Experiment 2 that it can even enhance performance in the difficult task involving reverse orientation

cubes (Figures 7.4 (left) & 7.5). In particular, reverse cubes shown in a ceiling perspective resulted in a large improvement. For the 18-item display, the improvement is around 100 msec.

It is well-known that texture density gradients can induce the percept of a receding plane (Gibson, 1950). We suggest that the textural mechanisms for the perception of surface slant are engaged here. When the reverse cubes are shown in perspective view, textural mechanisms for extracting ground-plane slant are driven by the apparent texture gradient. The resulting percept of a surface in 3-D would enhance the interpretation of the display as a physical scene, allowing the patterns to be perceived as 3-D shapes. Subjects' reports confirm that, indeed, the ceiling perspective enhanced the perception of the patterns as cubes hanging from the ceiling. For the floor perspective condition, some subjects described that the reverse orientation cubes looked like cubes balanced on a vertex.

In Experiment 3, the effect of contextual information is extended to a stimulus other than the shaded cube. We find that a consistent wall frame improves performance while an inconsistent one does not. We suggest that textural effects may be at work here also. Textural density has been shown to be an important factor in texture segmentation experiments (Julesz, 1981). For the consistent-frame experiment, the upper-right corner of the wall has the exact same configuration as that of the target (Figure 7.6A). This corner junction adds to the textural density of the target pattern, and may increase its saliency, consequently improving performance.

7.3.2 Expectation Effects

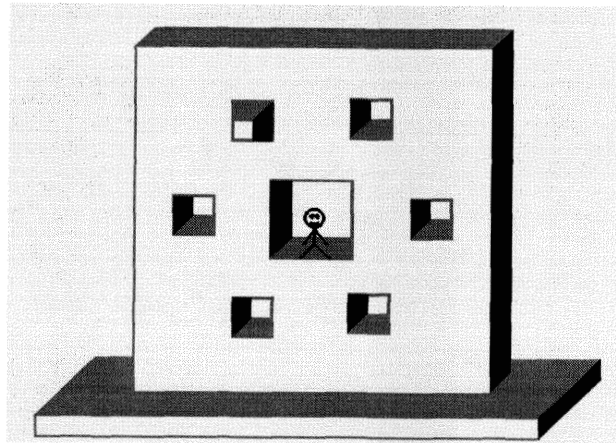
Textural mechanisms, however, cannot account for all of our observed contextual effects.

Evidence from Experiments 1 & 2 points to an influence other than textural effects. Adding a background room context gave strong effects in all cases (Figures 7.2 (right), 7.3, & 7.4 (right)). In Figure 7.2, we see that while perspective cues alone counteracted somewhat the effect of differential sizing, the background room context actually led to significant improvement for some display sizes. Similarly, in Experiment 2, the improvement went from being insignificant for all but the smallest display size to being statistically significant for every display size when the room background was added (Figure 7.4). Improvements for the 18- and 24-item displays were particularly large.

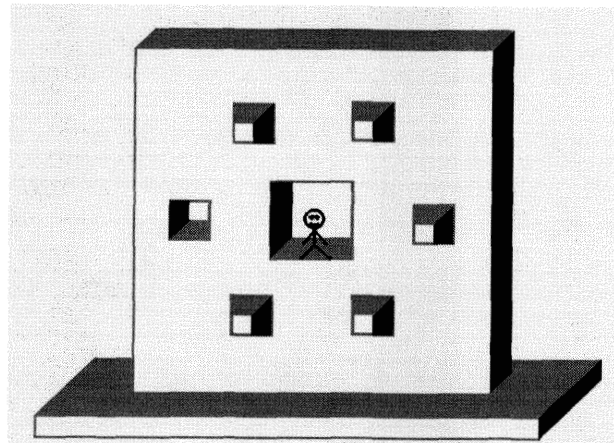
Can these effects be explained also by bottom-up textural mechanisms? Perhaps the two shaded junctions that made up the back corners of the room provided additional textural density? This is unlikely because these junctions do not have the same configuration or shading as either the distractors or the target. Perhaps the luminance discontinuity formed by the top or the base of the two side walls, depending on the experiment, served as "guidelines" for perceiving the texture gradient, thereby facilitating the texture gradient mechanisms for extracting surface slant? Again, we believe this effect to be minimal at most, since these lines are quite short, spanning less than a quarter of the height of the display.

We suggest, as an alternative, top-down expectation effects. Both the background room context in Experiments 1 & 2 and the wall frame in Experiment 3 were statically displayed on the screen and did not flash on and off at short durations with the target and distractor stimuli. These static background displays may have served as a constant reminder that the stimuli about to be flashed on should be given a particular 3-D interpretation. When the flashed stimuli were consistent with the preconceived scene interpretation, perception was facilitated, and performance was

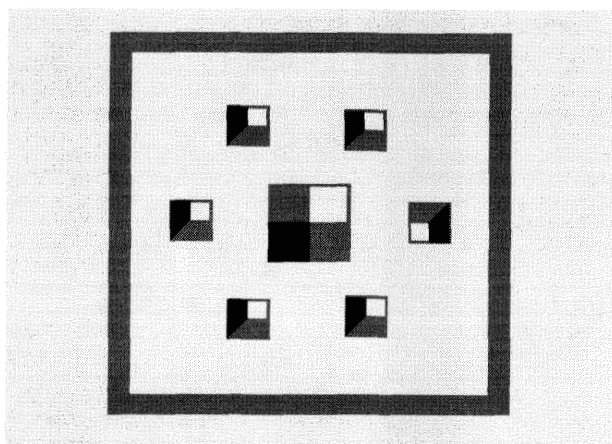
enhanced (Figures 7.2 (right), 7.4 (right), & 7.7 (left)). On the other hand, when the flashed stimuli were inconsistent with the preconceived interpretation, performance was impaired (Figure 7.3). This expectation effect may be related to the idea of top-down guidance in the "guided search model" proposed by Wolfe and collaborators. In the guided search model, attention can be guided in parallel by top-down information, allowing for increased search efficiency (Cave & Wolfe, 1990; Wolfe, Cave, & Franzel, 1989).



A



B



C

Figure 7.6: Consistent distractor holes in 3-D wall (A), inconsistent distractor holes in 3-D wall (B), and frame with no 3-D interpretation for control experiments (C)

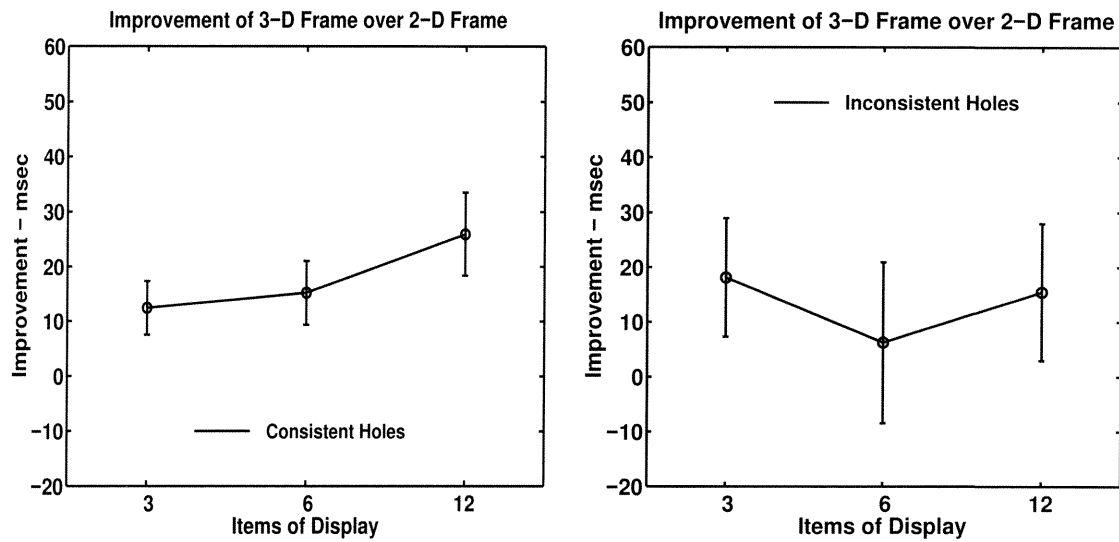


Figure 7.7: Improvement of performance for the consistent frame condition over the control condition is shown at left, and improvement of performance for the inconsistent frame condition is shown at right.

Chapter 8 Conclusion

8.1 Summary of Results

In the previous chapters, we have described experiments dealing with issues of both perception and mechanisms in the study of 3-D shape from shading. We have measured subjective shape perception from shaded figures and found the following (For details see Chapter 2):

1. Subjects display good precision in both tilt and slant measurements, contrary to previous findings which showed that slant consistency was poor. This high degree of consistency suggests that a robust shape percept is formed, probably supported by a priori shape processing models that are themselves robust and stable.
2. Shape percept is stable over relatively large variations of the reflectance function, further attesting to the robustness of the process.
3. Perception of surface normal is biased towards the normal to nearby tangent-plane discontinuities. This suggest that there is some robust, low-level process that assumes wrongly that such discontinuities are always occlusion boundaries.
4. Smaller shapes of 2 degrees or less are seen as more shapely than larger shapes, suggesting that there may be some stimulus size that optimally drives the shape mechanism.

Pertaining to the preattentive processes that subserve our perception of 3-D shape from shading, we summarize our results as follows:

1. Our 2AFC short duration SOA experiments confirm Enns & Rensink's finding that three dimensional shape from shading can be processed in parallel (See Chapter 3). We believe that this fast and parallel processing is dependent upon 3-D information because:
 - (a) Shaded stimuli that are easily interpretable as familiar three dimensional shapes are processed fast and in parallel while similar control stimuli that do not have such interpretation are not.
 - (b) Distractor-target reversal experiments that are equivalent in two-dimensional space, differing only in their 3-D interpretations, show asymmetry in performance. This asymmetry is seen with the cubes as well as the Y-junction in circles.
 - (c) 3-D contextual information can influence performance, both positively and negatively, depending on the degree to which the context contributes to a consistent 3-D interpretation, as suggested by the results of the experiments in Chapter 7.
 - (d) Subjects' reports of 3-D perception coincide with performance that indicates fast, parallel processing.
2. Our experiments suggest that this parallel 3-D process has the following characteristics:

- (a) *Has fast processing times*

For the normal orientation shaded cubes and pies, our experiments yielded

necessary SOA durations between 30 and 80 msec. These results suggest fast processing times for shaded 3-D stimuli, comparable to the ones previously reported in the classical "pop-out" and texture segregation experiments conducted using 2-D stimuli (Bergen & Julesz, 1983, Kröse, 1987, Gurnsey & Browse, 1987, Nothdurft, 1991).

(b) *Prefers shaded stimuli*

Unlike the results reported by Enns & Rensink (1991), our results indicate that unshaded line stimuli do not drive this fast and parallel process. They are processed more slowly and more serially. Other experimental results also support our finding that shading is a crucial component for 3-D pop-out; shaded bubbles, which contain no internal line edges, are found to be processed in parallel also (Braun 1990, 1993).

(c) *Computes locally on the Y-junction*

The normal orientation shaded Y-junction is a salient cue recognized by this 3-D process. Results from Experiment 3 & 5 of Chapter 3 suggest that computation begins locally at the Y-junction, and that perception of a complete 3-D solid is not necessary.

(d) *Subserves familiar shapes*

Familiar shapes in generic views drive this process better than unfamiliar ones. This is evidenced by the asymmetry experiments in Experiment 4, as well as Experiment 3 of Chapter 3. Convex, top-lit shapes, are processed with ease, while concave or bottom-lit shapes are not. Generic views of familiar shapes, such as the cube, are preferred. A similar positive effect of familiarity on search tasks concerning 2D line patterns was reported

recently by Wang & Cavanagh (1994). They also found that performance is better when the distractors are familiar than when the targets are.

(e) *Is closely tied to reflectance computations*

Our results from Chapter 4 suggest that early vision processes compute some aspects of apparent reflectance together with 3-D shape. Fast and parallel discriminations are better explained on the basis of apparent reflectance than that of perceptual 3-D shape or brightness.

(f) *Prefers top-left lighting*

Early vision 3-D mechanisms make *a priori* assumptions about direction of lighting that are not discounted even when over-all shading patterns and stereo cues indicate the contrary. Lighting-from-above is preferred over lighting from below, and our results thus far suggest that lighting-from-left is preferred over lighting-from-right (See Chapter 5).

(g) *Prefers convex shapes*

Early vision 3-D mechanisms also have a preference for convex over concave shapes (See Chapter 6). This preference does not appear to be overruled even when concave shapes are conferred using stereo disparity cues.

(h) *Is influenced by contextual information*

Our second set of experiments show that this 3-D process can be influenced by contextual information. Consistent contextual information that enhances the perception of a 3-D scene facilitates this process and improves performance, and inconsistent contextual information can impair performance. We suggest that these influences are mediated by bottom-up textural mechanisms as well as top-down expectation effects (For details

see Chapter 7.

8.2 The Biology of Shape from Shading

We would like to relate the scheme of 3-D shape processing to the existing literature on visual areas and pathways.

- Before 3-D shape can be computed, the separate components of shading information must first be processed. These components, or “ingredients” of shape fall under three general categories:

1. Shading Information

Shading, or luminance, information can be detected during the very first stages of the visual hierarchy. On-center and off-center cells of the LGN as well as the retina are sensitive to luminance information. Shading gradients, on the other hand, are more complex, and may be processed by cells higher up in the visual hierarchy. The receptive field sizes of visual cells increase as one proceeds up the processing hierarchy. In V1 and V2, cells have receptive fields of about 1 degree of visual angle. In Chapter 2 we reported the experimental result that “shapeliness” was best perceived when the shaded bubbles subtended about 1 degree of visual angle. This finding suggests that these shading gradients are being processed by V1, V2 cells. Furthermore, electrophysiological studies show that cells in the anterior inferotemporal cortex respond preferentially to shading gradients (Fujita *et al.*, 1992).

2. Boundaries and Contours

Oriented edge detector cells in V1 make good candidates for cells that

process boundary contour information. Hubel & Wiesel (1968) found cells in V1 that are specialized for detecting bars of a particular orientation. Furthermore, cells that respond to illusory contours have been found in V2 (von der Heydt *et al.*, 1984; von der Heydt & Peterhans, 1989). This finding lends support to the notion that percepts traditionally considered as the result of high-level cognitive processing, like shape from shading, can have neuronal correlates that reside relatively low on the visual processing hierarchy.

3. Corner Junctions

Corner junctions, like Y, T, and L-junctions, are important in shape from shading for defining boundary contours as well as occlusion boundaries. In general, they are necessary for figure-ground segmentation. One possibility is that end-stopping, or hypercomplex, cells in V1 may mediate the processing of corner junctions. These cells have inhibitory zones at one or both ends of the central excitatory zone (Hubel & Wiesel, 1962, 1968), and would therefore be appropriate for defining where a region ends. Simple and complex cells with end-inhibition have also been found (Dreher, 1972). By the same token, these cells may also contribute to the processing of corner junctions.

- These basic “ingredients” of shading information are combined to give different aspects of 3-D shape:

1. Depth

Large percentage of cells in V1 and V2 have been reported to be sensitive to stereo disparity (Poggio & Talbot, 1981; Poggio, 1984). There are

stereo cells that are tuned to a narrow range of stereo disparities, thereby encoding relatively specific depth information, and others that are tuned to a large range of disparities, encoding the more general properties of “far” and “near” depths. To date, there are however no physiological data showing cells that are sensitive to depth from shading cues.

2. Surface Orientation and Curvature

Gallant *et al.* reported in 1993 that many cells in V4 responded preferentially to polar or hyperbolic periodic stimuli over the traditional Cartesian sinusoidal gratings. It is hypothesized that these cells may be involved in the representation of surface shape.

3. Reflectance

V4 is the visual area that specializes in color processing (Zeki, 1973). We report in Chapter 4 that apparent reflectance is processed by early vision, in conjunction with 3-D shape. We hypothesize that luminance differences are discounted as being due to shading when a 3-D interpretation can be established. This is similar to the effect of “color constancy”. The large receptive field size of the cells in V4 and their chromatically opponent organization are thought to be properties that may contribute to the establishment of color constancy (Zeki, 1983). We suggest that 3-D shape mechanisms mediated in other visual areas may feed back onto Area V4, adjusting gain controls to discount luminance differences, resulting in a percept of constant reflectance. The fact that V4 forms the major input to the posterior and anterior inferotemporal cortices, areas which are strongly related to object recognition (Mishkin, 1982), supports also Area V4’s involvement in shape perception.

- In the processing of a 3-D scene, aspects of local shape, such as depth, surface

normals, and apparent reflectance, must eventually be integrated over wide areas of the visual field. We suggest that the final stages of visual processing take place in the posterior and anterior inferotemporal cortices (IT). Columns of cells that respond to very specific complex visual patterns, including shaded patterns, have been found in the anterior IT (Tanaka *et al.*, 1991; Fujita *et al.*, 1992). Lesion studies also implicate these areas in the process of object recognition (Mishkin, 1982; Mishkin *et al.*, 1983; Damasio, 1985).

Figure 8.1 summarizes schematically what our findings suggest regarding the processing of 3-D shapes in early vision.

Summary Diagram

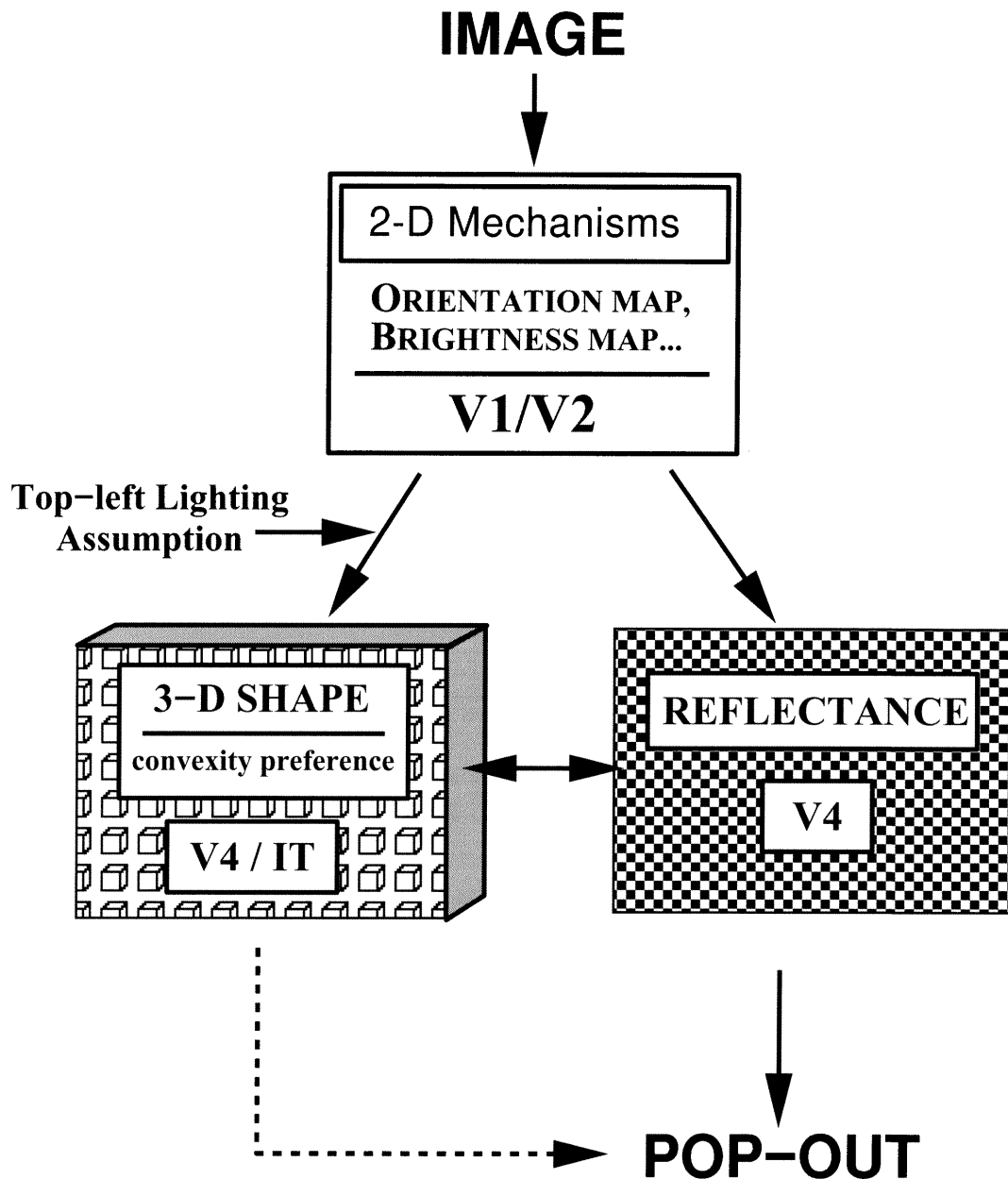


Figure 8.1: A schematic summarizing our findings concerning the processing of 3-D shapes by early vision processes.

Chapter 9 Appendix

9.1 A - The Surface Function

The function that describes our surfaces is $f(x, y) = f_1(x, y)f_2(x, y)$ where

$$f_1(x, y) = s\sqrt{1 - (x^2 + y^2)} + b_s(1 - (x^2 + y^2)^{\kappa/2})^{1/\kappa}$$

$$f_2(x, y) = k + ax + by + cx^2 + dy^2 + exy + g(|fx + gy + h + r(x * x + y * y)|)$$

and where

$$\kappa = 1.5$$

$$g(z) = m|z| + \gamma h(-t^2 z^2) + \delta h(-t|z|)$$

$$h(y) = \begin{cases} (1 - |y|)^2 & \text{if } -1 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Different choices of the parameters $s, b_s, k, a, b, c, d, e, f, g, h, r, m, \gamma, \delta, t$ produce different surfaces. The g function, driven by the γ, δ parameters, defines a bump of width $1/t$ with or without a dent on top.

The parameter values defining the different surfaces used are:

surface	s	b_s	k	a	b	c	d	e	f	g	h	r	m	γ	δ	t
1	1	0	1	0.8	0.2	0	0	0	0	0	0	0	1	0	0	1
2	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1
3	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	1
4	1	0	1	-0.2	0.1	0	0	0	0.3	0.3	-0.1	0	1	0	0	1
5	1	0	1	0	0	0	0	0	0.4	0.7	0.01	0	1	0	0	1
6	1	0	1	0	0	0	0	0	0.4	0.7	0.01	0.3	1	0	0	1

9.2 B - Method of Estimating Depth Scaling

Given the true surface equation $f(x, y)$, assume the perceived surface is a depth scaled version, i.e.; $sf(x, y)$, where the scaling parameter s is determined as the argument which minimizes the sum of norms of the differences between the normal vectors of the scaled surface and the measured normal vectors. That is,

$$s = \operatorname{argmin} \sum_i \|\vec{n}(x_i, y_i, s) - \vec{n}_m(x_i, y_i)\|$$

where i ranges over the set of measurements, $\vec{n}_m(x_i, y_i)$ is the mean normal vector measured at position (x_i, y_i) , and $\vec{n}(x_i, y_i, s)$ is the normal vector of the scaled surface.

9.3 C - Method of Estimating Psychometric Curves

Given the set of psychometric measurements $\{x_i, y_i, \sigma_i\}$, a psychometric curve is fitted with the equation $y = \operatorname{erf}((x - \mu)/\sigma)$. The parameters μ, σ are determined by solving

the nonlinear least squares problem in which the error vector \vec{e} has i^{th} component

$$\vec{e}_i = \frac{(y_i - erf((x_i - \mu)/\sigma))^2}{\sigma_i^2}$$

The covariance of the parameters $\{\mu, \sigma\}$ is also estimated according to

$$R_{\mu\sigma} = LRL^T$$

where

$$L = \begin{bmatrix} \frac{\partial^2 E}{\partial \mu^2} & \frac{\partial^2 E}{\partial \mu \partial \sigma} \\ \frac{\partial^2 E}{\partial \mu \partial \sigma} & \frac{\partial^2 E}{\partial \sigma^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial^2 E}{\partial \mu \partial y_1} & \cdots & \frac{\partial^2 E}{\partial \mu \partial y_N} \\ \frac{\partial^2 E}{\partial \sigma \partial y_1} & \cdots & \frac{\partial^2 E}{\partial \sigma \partial y_N} \end{bmatrix}$$

$$R = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_N \end{bmatrix}$$

$$E = \vec{e}\vec{e}^T$$

The μ parameter of the estimated psychometric curve is an estimate of the 50% level of perceived shapeliness for that given stimuli. The parameter $R_{\mu\sigma}$ provides an estimate of the standard deviation of μ , $\sigma_\mu = \sqrt{R_{\mu\sigma[1,1]}}$. Thus a weighted linear least squares relationship can be determined between stimuli size r_i and perceived shapeliness (μ_i, σ_{μ_i}) . The variance of the slope of this linear fit can also be estimated

according to

$$\sigma_{slope}^2 = \frac{\sum_i \frac{r_i^2}{\sigma_{\mu_i}^2}}{\sum_i \frac{1}{\sigma_{\mu_i}^2} \sum_i \frac{r_i^2}{\sigma_{\mu_i}^2} - \left(\sum_i \frac{r_i}{\sigma_{\mu_i}^2} \right)^2}$$

Bibliography

- [1] Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science*, *262*, 2042-2044.
- [2] Beck, J. (1966). Effects of orientation and of shape similarity on perceptual grouping. *Perception & Psychophysics*, *1*, 300-302.
- [3] Beck, J. (1967). Perceptual grouping produced by line figures. *Perception & Psychophysics*, *2*, 491-495.
- [4] Beck, J. (1982). Textural segmentation. In J. Beck (Ed.), *Organization and Representation in Perception*, pp. 285-317, Hillsdale, NJ: Lawrence Erlbaum.
- [5] Bergen, J. R. & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature*, *303*(5919), 696-698.
- [6] Bergstrom, S.S. (1977) Common and relative components of reflected light as information about the illumination, colour, and three-dimensional form of objects. *Scandinavian Journal of Psychology*, *18*, 180-186.
- [7] Braun, J. (1990). Focal attention and shape-from-shading. *Perception*, *19*, A112.
- [8] Braun, J. (1993). Shape-from-shading is independent of visual attention and may be a 'texton'. *Spatial Vision*, *7*(4), 311-322.
- [9] Braunstein, M. L., Andersen, G. J., & Riefer, D. M. (1982). The use of occlusion to resolve ambiguity in parallel projections. *Perceptual Psychophysics*, *31*, 261-267.

- [10] Brewster, D. (1847). On the conversion of relief by inverted vision. *Edinburgh Philosophical Transactions*, 15, 657.
- [11] Bulthoff, H. H., & Mallott, H. A. (1992). Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America*, A5, 1749-1758.
- [12] Cave, K. R. & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, 22, 225-271.
- [13] Cutting, J. E. & Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal Experimental Psychology*, 113, 198-216.
- [14] Damasio, A. R. (1985). Prosopagnosia. *Trends in Neuroscience*, 8, 132-135.
- [15] Dreher, B. (1972). Hypercomplex cells in the cat's striate cortex. *Investigative Ophthalmology*, 11, 355-356.
- [16] Duncan, J. & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3), 433-458.
- [17] Enns, J. T. & Rensink, R. A. (1990). Influence of scene-based properties on visual search. *Science*, 247, 71-723.
- [18] Enns, J. T. & Rensink, R. A. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review*, 98(3), 335-351.
- [19] Fujita, I., Tanaka, K., Ito, M., & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, 360, 343-346.
- [20] Gallant, J. L., Braun, J., & Van Essen, D. C. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science*, 259, 100-

- [21] Gibson, J. (1950). *The Perception of the Visual World*, Boston, MA: Houghton Mifflin.
- [22] Gilchrist, A. L. (1979). The perception of surface blacks and whites. *Scientific American*, 240(3), 112-124.
- [23] Gombrich, E. H. (1976). *Gombrich on the Renaissance*, 3. London, England: Phaidon.
- [24] Green, M. (1991). Visual search, visual streams, and visual architectures. *Perception & Psychophysics*, 50(4), 388-403.
- [25] Gurnsey, R. & Browse R. A. (1987). Micropattern properties and presentation conditions influencing visual texture discrimination. *Perception and Psychophysics*, 40(3), 239-252.
- [26] He, Z. J. & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature*, 359(6392), 231-233.
- [27] He, Z. J. & Nakayama, K. (1993). Common surface rather than common depth determines attention in 3-D search task. *Society for Neuroscience Abstracts*, 19(1), 773.
- [28] Horn, B. (1975). Obtaining shape from shading information. In P.H. Winston (Ed.), *Psychology of computer vision*. New York: McGraw-Hill.
- [29] Hubel, d. H. & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160, 106-154.

- [30] Hubel, d. H. & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 225,, 215-243.
- [31] Julesz, B. (1975). Binocular depth perception without familiarity cues. *Science*, 145, 356-362.
- [32] Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature*, 290, 91-97.
- [33] Julesz, B. (1984). A brief outline of the texton theory of human vision. *Trends in Neurosciences*, 7(2), 41-45.
- [34] Julesz, B. (1984). Toward an axiomatic theory of preattentive vision. In G. M. Edelman, W. E. Gall & M. Cowan (Eds.), *Dynamic Aspects of Neocortical Function*, pp. 585-611. New York: Neurosciences Research Foundation.
- [35] Julesz, B. (1986). Texton gradients: the texton theory revisited. *Biological Cybernetics*, 54, 245-251.
- [36] Kleffner, D. A. & Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception & Psychophysics*, 52(1), 18-36.
- [37] Knill, D. C., & Kersten, D. (1991). Apparent surface curvature affects lightness perception. *Nature*, 351, 228-230.
- [38] Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. L. (1992). Surface perception in pictures. *Perception & Psychophysics*, 52(5), 487-496.
- [39] Kröse, B. J. A. (1987). Local structure analyzers as determinants of preattentive pattern discrimination. *Biological Cybernetics*, 55, 289-298.

- [40] Lappin, J. S., Doner, J. F., & Kottas, B. (1980). Minimal conditions for the visual detection of structure and motion in three dimensions. *Science*, 209, 717-719.
- [41] McNichol, D. (1972). *A Primer of Signal Detection Theory*, London, UK: George Allen & Unwin.
- [42] Neisser, U. (1967). *Cognitive psychology*, Englewood Cliffs, NJ: Prentice-Hall.
- [43] Metelli, F. (1974). The perception of transparency. *Scientific American*, 230(4), 91-98.
- [44] Mingolla, E., & Todd., J. T. (1989). *Shape from shading*. Cambridge, MA: MIT.
- [45] Mishkin, M. (1972). Cortical visual areas and their interactions. In A. G. Karczmar & j. C. Eccles (Eds.), *Brain and Human Behavior*, Berlin: Springer-Verlag.
- [46] Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two critical pathways. *Trends in Neuroscience*, 6, 414-417.
- [47] Nakayama, K. & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, 257(5075), 1357-1363.
- [48] Nakayama, K. & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, 320(6059), 264-265.
- [49] Nothdurft, H. C. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Research*, 31(6), 1073-1078.
- [50] Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9, 97-113.

- [51] Olson, R. & Attneave, F. (1970). What variables produce similarity grouping? *American Journal of Psychology*, 83, 1-21.
- [52] Pentland, A. P. (1982). Finding the direction of illumination. *Journal of the Optical Society of America*, 72, 448-455.
- [53] Poggio, G. F. (1984). Processing of stereoscopic information in monkey visual cortex. In G. M. edelman, W. E. Gall & W. M. Cowan (Eds.), *Dynamic Aspects of Neocortical Function*, New York: Wiley.
- [54] Poggio, G. F. & Talbot, W. H. (1981). Mechanism of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *Journal of Physiology*, 315, 469-492.
- [55] Ramachandran, V. S. (1988). Perception of shape from shading. *Nature*, 331(6152), 163-166.
- [56] Rasmussen, T. & Milner, B. (1977). The role of early left-brain injury in determining lateralization of cerebral speech functions. *N.Y. Academy of Science*, 299, 355-379.
- [57] Rittenhouse, D. (1786). Explanation of an optical deception. *Transactions of the American Philosophical Society*, 2, 37-42.
- [58] Rubenstein, B. & Sagi, D. (1990). Spatial variability as a limiting factor in texture-discrimination tasks – implications for performance asymmetries. *J. Opt. Soc. Am. A*, 7(9), 1632-1643.
- [59] Sagi, D., & Julesz, B. (1985). “Where” and “what” in vision. *Science*, 228, 1217-1219.

- [60] Stevens, K. A. (1983a). Slant-tilt: The visual encoding of surface orientation. *Biological Cybernetics*, 53, 137-151.
- [61] Stevens, K. A. (1983b). Surface tilt (the direction of slant): A neglected psychophysical variable. *Perception & Psychophysics*, 33, 241-250.
- [62] Sun, J. Y., Mennucci, A.C., Goncalves, L.P., & Perona, P. (1994). Contour and scale influence on shape perception from shading. *CIT/CNS-TR-94-03 California Institute of Technology*, 1-27.
- [63] Sun, J. Y. & Perona, P. (1995). Influence of 3-D shape interpretation on preattentive popout and 2-D mechanisms *Investigative Ophthalmology & Visual Science*, 36(4), 902.
- [64] Sun, J. Y. & Perona, P. (1996a). Early Computation of Shape and Reflectance in the Visual System *Nature*, 379, 165-168.
- [65] Sun, J. Y. & Perona, P. (1996b). Where is the Sun? *Investigative Ophthalmology & Visual Science*, 37(4), 4283.
- [66] Sun, J. Y. & Perona, P. (1996c). Preattentive perception of elementary three dimensional shapes. *Vision Research*, In press.
- [67] Symons, L. A. , Humphrey, G. K., Herbert, A. M., & Goodale, M. A. (1993). Region segregation in a visual form agnostic. *Investigative Ophthalmology & Visual Science*, 34(4), 1084.
- [68] Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology*, 66, 170-189.

- [69] Todd, J. T., & Akerstrom, R. A. (1987). Perception of three-dimensional form from patterns of optical texture. *Journal of Experimental Psychology: Human Perception & Performance*, 13, 242-255.
- [70] Todd, J. T., & Mingolla, E. (1983). Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception & Performance*, 9(4), 583-595.
- [71] Todd, J. T., & Reichel, F. D. (1989). Ordinal structure in the visual perception and cognition of smooth surfaces. *Psychological Review*, 96, 643-657.
- [72] Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 194-214.
- [73] Treisman, A. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97-136.
- [74] Treisman, A. & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95, 15-48.
- [75] Treisman, A. & Patterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experiment Psychology: Human Perception and Performance*, 10, 12-31.
- [76] Treisman, A. & Southner, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, 114, 285-310.

- [77] von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984) Illusory contours and cortical neuron responses. *Science*, *224*, 1260-1262.
- [78] von der Heydt, R., & Peterhans, E. (1989a) Mechanisms of contour perception in monkey visual cortex. I. *Journal of Neuroscience*, *9*, 1731-1748.
- [79] Wang, Q. Q., Cavanagh, P. & Green, M. (1994). Familiarity and popout in visual-search. *Perception* *23* *Psychophysics*, *56*(5), 495-500.
- [80] Williams, D. & Julesz, B. (1992). Perceptual asymmetry in texture perception. *Proc. Natl. Acad. Sci. USA*, *89*, 6531-6534.
- [81] Wolfe, J. M., Cave, K. R. & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model of visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 419-433.
- [82] Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I. & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 34-49.
- [83] Wolfe, J. M. & Franzel, S. L. (1988). Binocularity and visual search. *Perception* *17* *Psychophysics*, *44*(1), 81-93.
- [84] Yonas, A., Kuskowski, M., & Sternfels, S. (1979). The role of frames of reference in the development of responsiveness to shading. *Child Development*, *50*, 495-500.
- [85] Zeki, S. M. (1973). Color coding in rhesus monkey prestriate cortex. *Brain Research*, *53*, 422-427.
- [86] Zeki, S. M. (1983). Color coding in the cerebral cortex: The reaction of cells in monkey visual cortex to wavelengths and colours. *Neuroscience*, *9*, 741-765.